

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 7, 2004, 11:41:13 ; Search time 1883 Seconds  
(without alignments)  
5478.302 Million cell updates/sec

Title: US-09-437-450a-40  
Perfect score: 238  
Sequence: 1 ttttttttttttgggagga.....tacgtggcaaaaaaaaaa 238

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: GenBank:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_ro.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rnd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htg\_hum.\*
- 40: em\_htg\_mus.\*
- 41: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	230.2	96.7	907	3	AF301606	AF301606 Caenorhab
2	194.4	81.7	23602	3	CEP3588	Z81529 Caenorhabdi
C 3	95.4	40.1	36974	3	CEK01D12	Z75543 Caenorhabdi
C 4	95.4	40.1	40450	3	CEC54D1C	Z75531 Caenorhabdi
C 5	41	17.2	88549	3	AC116979	AC116979 Dictyoste
C 6	41	17.2	193057	9	AC022325	AC022325 Homo sapi
C 7	40.2	16.9	138683	2	AC108193	AC108193 Felis cat
C 8	39.2	16.5	250470	2	AC126197	AC126197 Rattus no
C 9	39	16.4	177864	2	AC146595	AC146595 Mus muscu
C 10	38.8	16.3	128635	9	AC108074	AC108074 Homo sapi
C 11	38.8	16.3	138772	2	AP004700	AP004700 Oryza sat
C 12	38.8	16.3	147655	8	AP005757	AP005757 Oryza sat
C 13	38.8	16.3	161010	2	AC132361	AC132361 Mus muscu
C 14	38.6	16.2	1182	3	AF147831	AF147831 Stenogeph
C 15	38.6	16.2	132389	2	AC108190	AC108190 Felis cat
C 16	38.6	16.2	132983	8	AC130799	AC130799 Medicago
C 17	38.5	16.2	144988	10	AC101022	AC101022 Mus muscu
C 18	38.6	16.2	146405	10	AC122795	AC122795 Mus muscu
C 19	38.6	16.2	193366	5	AL935279	AL935279 Zebrafish
C 20	38.2	16.1	71032	2	AC016983	AC016983 Mus muscu
C 21	38.2	16.1	170934	2	AC118982	AC118982 Gallus ga
C 22	38.2	16.1	197925	10	AL672241	AL672241 Mouse DNA
C 23	38.2	16.1	230352	2	AC016982	AC016982 Mus muscu
C 24	38.2	16.1	251891	2	AC084742	AC084742 Mus muscu
C 25	38	16.0	184323	8	AP005066	AP005066 Oryza sat
C 26	37.6	15.8	68555	2	AC104963	AC104963 Homo sapi
C 27	37.6	15.8	79509	9	AL353769	AL353769 Human DNA
C 28	37.6	15.8	154158	2	AL365257	AL365257 Homo sapi
C 29	37.6	15.8	166847	9	CNS06C7S	AL390801 Human chr
C 30	37.6	15.8	176505	9	CNS06C7O	AL390335 Lemur chr
C 31	37.4	15.7	203231	9	AC123543	AC123543 Lemur chr
C 32	37.2	15.6	159462	2	AC120991	AC120991 Oryza sat
C 33	37	15.5	64096	9	AL138727	AL138727 Human DNA
C 34	37	15.5	169713	2	AC016733	AC016733 Homo sapi
C 35	36.6	15.4	176146	2	AC023124	AC023124 Homo sapi
C 36	36.6	15.4	212233	10	AL772216	AL772216 Mouse DNA
C 37	36.4	15.3	164288	2	AC128949	AC128949 Rattus no
C 38	36.4	15.3	182527	2	AC117010	AC117010 Rattus no
C 39	36.2	15.2	4233	8	AF038045	AF038045 Gossypium
C 40	36.2	15.2	112427	8	AC122164	AC122164 Medicago
C 41	36	15.1	167069	9	CNS06C88	AL391516 Human chr
C 42	36	15.1	204444	2	AL445436	AL445436 Homo sapi
C 43	36	15.1	226711	2	AC127440	AC127440 Rattus no
C 44	35.8	15.0	11542	1	AE010610	AE010610 Fusobacte
C 45	35.8	15.0	116392	9	HS324C6	AL031586 Human DNA

ALIGNMENTS

RESULT 1  
AF301606  
LOCUS  
DEFINITION  
Caenorhabditis elegans cadmium-inducible lysosomal protein CDR-1  
907 bp mRNA linear INV 28-OCT-2002

ACCESSION  
AF301606  
VERSION  
AF301606.1  
KEYWORDS  
AF301606.1  
SOURCE  
Caenorhabditis elegans  
ORGANISM  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE  
1 (sites)  
AUTHORS  
Liao V.H. and Freedman J.H.  
TITLE  
Cadmium-regulated genes from the nematode Caenorhabditis elegans.

Identification and cloning of new cadmium-responsive genes by differential display  
 J. Biol. Chem. 273 (48), 31962-31970 (1998)  
 99241962  
 9622667  
 2 (bases 1 to 907)  
 Liao, V.H.C., Dong, J. and Freedman, C.H.  
 Molecular Characterization of a Novel, Cadmium-inducible Gene from the Nematode *Caenorhabditis elegans*. A NEW GENE THAT CONTRIBUTES TO THE RESISTANCE TO CADMIUM TOXICITY  
 J. Biol. Chem. 277 (44), 42049-42059 (2002)  
 12189149  
 3 (bases 1 to 907)  
 Freedman, J.H. and Liao, V.H.-C.  
 Direct Submission  
 Submitted (30-AUG-2000) Nicholas School of the Environment, Duke University, Box 90328, Durham, NC 27708-0328, USA  
 Derived from cosmid F35E8.  
 Location/Qualifiers  
 1..907  
 /organism="Caenorhabditis elegans"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:6239"  
 /note="derived from cosmid F35E8"  
 1..907  
 /gene="cdr-1"  
 1..15  
 /gene="cdr-1"  
 16..849  
 /gene="cdr-1"  
 /codon\_start=1  
 /evidence=experimental  
 /product="cadmium-inducible lysosomal protein CDR-1"  
 /protein\_id="AAG30870.1"  
 /db\_xref="GI:11120429"  
 /trans="atons="WDSCILITTAALPGAVLYLKNFTVPSIKPKEDLHKDYKKOV  
 VTLQMRKNCNENSPGCKMIEILCEFKIPYELLCTTSRSGNLVFPVLENGEHI  
 ADSLLEIRLSPKIPSPTELBQTSVALSFADHLLFVLLIRKLVDFEYKILIE  
 IIGTFINFLMLPLKAIIGKNYKNCQAGDPELSELDLHRLRIVENTLAKK  
 KFLGEBITAAATVPSGLATVYFFRNHISDLEKDFPKLLEYCERVRHEVYPKDFT  
 Y"  
 850..907  
 /gene="cdr-1"  
 878..883  
 /gene="cdr-1"  
 /evidence=experimental  
 polyA\_signal  
 ORIGIN  
 Query Match 96.7%; Score 230.2; DB 3; Length 907;  
 Best Local Similarity 98.7%; Pred. No. 3.7e-50;  
 Matches 232; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 TTTTCTTTTGGAGAAATCAGCGCGGATCGACAGCTCTTCTCAATTGGCAACT 63  
 DB 673 TTTCTTTTGGGAGAAATCAGCGCGGATCGACAGCTCTTCTCAATTGGCAACT 732  
 QY 64 GTCTATATCCATTCGCAATCACATTTGGACGTTCTCGAAAGAGATTCGCCAAGTTA 123  
 DB 733 GTCTATATCCATTCGCAATCACATTTGGATGTTCTCGAAAGAGATTCGCCAAGTTA 792  
 QY 124 TTGAGTACTGTGAAGAGTTCGTATGAGTTTACCCAGAGCTTTACTATGTAATT 183  
 DB 793 TTGGAGTACTGTGAAGAGTTCGTATGAGTTTACCCAGAGCTTTACTATGTAATT 852  
 QY 184 AAA-TGTCAACTAGTAGTCAGATCAATTAATTTCTACGTGGCAAAAAA 238  
 DB 853 AATTGTCAACTAGTAGTCAGATCAATTAATTTCTACGTGGCAAAAAA 907  
 RESULT 2  
 CEF35E8 23602 bp DNA linear INV 03-DEC-2003  
 LOCUS  
 DEFINITION *Caenorhabditis elegans* cosmid F35E8, complete sequence.

ACCESSION Z81529  
 VERSION Z81529.1 GI:2653095  
 KEYWORDS HTG.  
 SOURCE *Caenorhabditis elegans*  
 ORGANISM *Caenorhabditis elegans*  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Filarioidea; *Caenorhabditis*.  
 REFERENCE 1 none.  
 AUTHORS Genome sequence of the nematode *C. elegans*: a platform for  
 TITLE investigating biology. The *C. elegans* Sequencing Consortium  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 39069613  
 PUBMED 9851916  
 REMARK The *C. elegans* Sequencing Consortium.  
 REFERENCE 2 (bases 1 to 23602)  
 AUTHORS Gardner A.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger  
 Institutes, Hinxton, Cambridge CB10 1SA, England and Department of  
 Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
 jesssanger.ac.uk or rwenematode.wustl.edu  
 On Nov 23, 1997 this sequence version replaced gi:2546904.  
 Coding sequences below are predicted from computer analysis, using  
 predictions from Genefinder (P. Green, U. Washington), and other  
 available information.  
 Current sequence finishing criteria for the *C. elegans* genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 For a graphical representation of this sequence and its analysis  
 see: - [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F35E8)  
 name=F35E8  
 IMPORTANT: This sequence is NOT necessarily the entire insert of  
 the specified clone. It may be shorter because we only sequence  
 overlapping sections once, or longer because we arrange for a small  
 overlap between neighbouring submissions.  
 IMPORTANT: This sequence is not the entire insert of clone F35E8.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.  
 The true left end of clone F35E8 is at 18616 in  
 sequence CEF06C12.  
 The true right end of clone F35E8 is at 23602 in this sequence. The  
 true left end of clone F05G11 is at 1324 in this sequence. The true  
 right end of clone F36C12 is at 104 in this sequence. The start of  
 this sequence (1..104) overlaps with the end of sequence CEF06C12.  
 The end of this sequence (23499..23602) overlaps with the start of  
 sequence CEF05G11.  
 FEATURES  
 source  
 1..23602  
 /organism="Caenorhabditis elegans"  
 /mol\_type="genomic DNA"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /clone="F35E8"  
 ORIGIN  
 Query Match 81.7%; Score 194.4; DB 3; Length 23602;  
 Best Local Similarity 99.5%; Pred. No. 7.4e-41;  
 Matches 195; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 30 GGCGATGCAACAGCTTCTCTCAATTGGCACTGCTATTATTCATCCGCAATCAAT 89  
 DB 18625 GGCGATGCAACAGCTTCTCTCAATTGGCACTGCTATTATTCATCCGCAATCAAT 18684  
 QY 90 TTGGATGTTCTCGAAAGAGCTTCCCAAGATTTATGGAGTCTGTCAAGAGTTCGTCA 149  
 DB 18685 TTGGATGTTCTCGAAAGAGCTTCCCAAGATTTATGGAGTCTGTCAAGAGTTCGTCA 18744  
 QY 150 TGAAGTTTACCCAAAGAGCTTTACTATGTAATTAATTTGCAACATGATGTCAGATCA 209



11987..12085,12159..12481,13776..13919,14219..14302}}  
 /gene="twk-14"  
 complement(join(111177..11444,11518..11592,11643..11801,  
 11987..12085,12159..12481,13776..13919,14219..14302))  
 /gene="twk-14"  
 /standard name="K01D12.4"  
 /note="similar to potassium channel protein  
 cDNA EST YK743d10.5 comes from this gene  
 cDNA EST YK743d10.3 comes from this gene"  
 /codon\_start=1  
 /product="C. elegans TWK-14 protein (corresponding  
 sequence K01D12.4)"  
 /protein\_id="CAA99871.1"  
 /db\_xref="GI:3878145"  
 /translation="WHKQFGIGQREMLRANTLPSITRAKVGCFARLRIEENARF  
 VLICILLYAFAGALLPHMLENEVDRIALDNRMADYQVCKHKLPCDFEEM  
 VRFISDGRGSLNRSRFDHLSGSPFSAIVSTIGFTSTPRHLGRITIVYGVWG  
 CTCWLFNLFLELVGTGSLVLSRERKIRYKESGKPKPTLLNNEDNESSSS  
 CGHMDNWRVSVYKVFILFSCMLVLTASAGIYVWENYIDSLYFCFISFATIG  
 GDVYSNQQDVTFRMSPDLYRFVFCLLTGACFFYCLSVSVIVVQLLMMIKMDVK  
 VEDRSFCFKKRYMGLGLRPPKIGISNLGAGNCCIEHLNVFRIRYDI"  
 join(21543..2227,22273..22403)  
 /gene="K01D12.5"  
 -oin(21543..2227,22273..22403)  
 /gene="K01D12.5"  
 /standard name="K01D12.5"  
 /note="cDNA EST OSTF168B6.1 comes from this gene  
 cDNA EST OSTR168B6.1 comes from this gene"  
 /codon\_start=1  
 /product="Hypothetical protein K01D12.5"  
 /protein\_id="CAA99872.2"  
 /db\_xref="GI:31414821"  
 /translation="MGRQSLALLSILIGETIANEYDGLSSHQSHQSVTSAPFHV  
 SEQQTNSRVNPOQOKESVLYQVLPESRVYQVYLYZTHGTYSNITQVGP  
 YQMGSAPGNSVYSGNESYGVNQAPVTPAPVHFORSYSAAPTPSPSE  
 TRENNVAPGHHAPQAQQHRSPPSPFPMIPKPVLSYGAMPFLANLPPAQGV  
 YGAPDSHSGYDGFSLAPAPPSSGSSVDVDTWALEASIDANTRRHRIF"  
 complement(join(22649..22801,22850..23341,23437..23753,  
 23798..24207,24280..24377,24423..24458))  
 /gene="K01D12.6"  
 complement(join(22649..22801,22850..23341,23437..23753,  
 23798..24207,24280..24377,24423..24458))  
 /gene="K01D12.6"  
 /standard name="K01D12.6"  
 /note="cDNA EST cm1496 comes from this gene  
 cDNA EST YK384f1.3 comes from this gene  
 cDNA EST YK384f1.5 comes from this gene  
 cDNA EST YK1046c01.5 comes from this gene  
 cDNA EST YK1046c01.3 comes from this gene  
 cDNA EST OSTF196H4.1 comes from this gene  
 cDNA EST OSTR196H4.2 comes from this gene  
 cDNA EST OSTR196H4.1 comes from this gene"  
 /codon\_start=1  
 /product="Hypothetical protein K01D12.6"  
 /protein\_id="CAA99873.3"  
 /db\_xref="GI:24817517"  
 /translation="MTSLYMLQFGLGINKNEIITKTVLSVAGIGIVFANFVRYW  
 ROKREKVPQSIIEIGKADIIIVPSRLRSQGRPPSPQSGERSLRSRQ  
 SQRNMPTEADITSGBELCTELRTKIVHNLEMYNRSSKDMERSIKTEGLKGLK  
 OVEDIVLLVPMEDFERDNNMFYVSGVAGSVTRGSRSTLSVLSQDPSRFAVER  
 FACIDIDDFVSDANLKNELRFLDSGQALNGEVYKRSRMEPCDSDSPAAK  
 LYCVKRALNLKDEHRVWLAKGRTLLADFIHRTKQDPVAFNADENLEVSNDK  
 NEEQLRQVREGSGVCTGYDAIDFIILDAFELKSPSPSVSVYTKNYPNSVMKYS

Query Match 40.1%; Score 95.4; DB 3; Length 36974;  
 Best Local Similarity 71.2%; Pred. No. 8e-15;  
 Matches 126; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 4 TTTTITTTTGGAGGAATCAGCGGGCGGATGCAACAGCTTCTCTCAATGGCACT 63  
 34856 TATTTATTTGGAGACCAATCTGCCAGTTGATGGAGCGTCTTTAGCCAGTGGCAGTT 34797  
 QY 64 GTCTATTATTCGATTCGGAATCACAATTCGGAGTGTCTCGAAGAGCACTCCCAAGTTA 123

Db 34796 GTTTATTATTCATTTTACATCACTCTACTGCTGTTGAAATGATTCACAGATT 34737  
 QY 124 TTGCAGTACTGTGAAGAGTTCGTCAATGAAGTTTACCCAAAGCACTTACTATGTA 180  
 Db 34736 CTACGATCTGTGAAGGATCCCGAAGGAATTTATCCAAATGATTTACTATTGA 34680

RESULT 4  
 CECS4D10/c 40450 bp DNA linear INV 03-DEC-2003  
 Caenorhabditis elegans cosmid C54D10, complete sequence.  
 ACCESSION Z75531  
 VERSION Z75531.1 GI:1418465  
 HTG: Serine protease inhibitor.  
 KEYWORDS Caenorhabditis elegans  
 SOURCE Caenorhabditis elegans  
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1  
 Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium  
 Science 282 (5396), 2012-2018 (1998)  
 99069613  
 9851916  
 The C. elegans Sequencing Consortium.  
 2 (bases 1 to 40450)  
 Direct Submission  
 Dobson, R.  
 TITLE Submitted (29-JUN-1996) Nematode Sequencing Project, Sanger  
 Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
 Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
 jess@sanger.ac.uk or wrne-atode.wustl.edu  
 Coding sequences below are predicted from computer analysis, using  
 predictions from Genefinder (P. Green, U. Washington), and other  
 available information.  
 Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 For a graphical representation of this sequence and its analysis  
 see: http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?  
 name=C54D10  
 IMPORTANT: This sequence is NOT necessarily the entire insert of  
 the specified clone. It may be shorter because we only sequence  
 overlapping sections once, or longer because we arrange for a small  
 overlap between neighbouring submissions.  
 IMPORTANT: This sequence is not the entire insert of clone C54D10.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.  
 The true left end of clone C54D10 is at 1 in this sequence. The  
 true right end of clone C54D10 is at 1020 in  
 sequence CEF28C1.  
 The true left end of clone F28C1 is at 40353 in this sequence. The  
 true right end of clone K01D12 is at 4214 in this sequence. The  
 start of this sequence (1..4214) overlaps with the end of sequence  
 CEK01D12.  
 The end of this sequence (40353..40450) overlaps with the start of  
 sequence CEF28C1.  
 FEATURES  
 source 1..40450  
 Location/Qualifiers  
 /organism="Caenorhabditis elegans"  
 /mol\_type="genomic DNA"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /clone="C54D10"

ORIGIN  
 Query Match 40.1%; Score 95.4; DB 3; Length 40450;  
 Best Local Similarity 71.2%; Pred. No. 8e-15;







```
Query Match      17.2%; Score 41; DB 3; Length 193057;
Best Local Similarity 52.0%; Pred. No. 1.4;
Matches 92; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 60 AACTGCTATTATCCATTCGCAACACATTTGGATGTTCTCGAAGAGACTTCCCAA 119
DB 52787 ACAGCCATCATATATGGGGATTTACCAATCCAAATACATTTAAAGCTATCTTAA 52728

QY 120 GTTATGAGTACTGTGMAAGAGTTCGTCATGAAGTTTACCCAAAGGACTTTACTATGTG 179
DB 52727 GTCAATGATTAAGCCGACATTTTTCGAAATACAGATTAACATTCCTAAAT 52668

QY 180 AATTAATGTCACAACTAGTAGTCAGATCAATAAATCTCTACGTGCAAAAAA 236
DB 52667 TATGAATTTACATCTTTTATGAAAGTACAAAAAGGAAAGGTTGAGAACCA 52611

RESULT 7
AC108193/c
LOCUS      Felis catus clone RP86-469M8, WORKING DRAFT SEQUENCE, 2 ordered
DEFINITION
ACCESSION  AC108193
VERSION     AC108193.2 GI:25815311
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE     Felis catus (cat)
ORGANISM   Felis catus
REFERENCE  1 (bases 1 to 138683)
AUTHORS   Akhtar, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
          Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
          Cariaga, K., Coleman, B., Engle, J., Granite, S., Guat, X., Gupta, J.,
          Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,
          Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
          Margulies, E.H., Masello, C., Maskeri, B., McDowell, J.,
          Pagnirilan, C., Pearson, R., Portnoy, M.E., Prasad, A.,
          Reddix-Bugue, A., Schandler, K., Schueler, M.G., Sison, C.,
          Stantripp, S., Thomas, J.W., Thomas, P.J., Torchman, J.W., Vogt, J.L.,
          Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.
          NISC Comparative Sequencing Initiative
          Unpublished
          2 (bases 1 to 138683)
          Green, E.D.
          Direct Submission
          Submitted (26-JAN-2002) NIH Intramural Sequencing Center, 8717
          GroveMont Circle, Gaithersburg, MD 20877, USA
          3 (bases 1 to 138683)
          Green, E.D.
          Direct Submission
          Submitted (28-NOV-2002) NIH Intramural Sequencing Center, 8717
          GroveMont Circle, Gaithersburg, MD 20877, USA
          On Nov 28, 2002 this sequence version replaced gi:18376892.
          ----- Genome Center
          Center: NIH Intramural Sequencing Center
          Center code: NISC
          Web site: http://www.nisc.nih.gov
          Contact: nisc.zoo@nhgri.nih.gov
          ----- Project Information
          Center project name: coe
          Center clone name: 469M08
```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated

with a Phrap-derived quality score.

```
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 138519 bases at least Q40
Consensus quality: 138564 bases at least Q30
Consensus quality: 138577 bases at least Q20
Insert size: 133000; agarose-fp
Insert size: 138583; sum-of-contigs
Quality coverage: 12.26x in Q20 bases; agarose-fp
Quality coverage: 11.76x in Q20 bases; sum-of-contigs
```

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 105452: contig of 105452 bp in length  
\* 105453 105552: gap of unknown length  
\* 105553 138683: contig of 33131 bp in length.

#### FEATURES

##### source

```
1. 138683
  /organism="Felis catus"
  /mol_type="genomic DNA"
  /db_xref="taxon:9685"
  /clone="RP86-469M8"
  /clone_lib="RP86"
```

##### misc\_feature

```
1. 105452
  /note="assembly_fragment"
  /clone_end:SP6
  /vector_side:left"
```

##### misc\_feature

```
1. 61312
  /note="clone overlaps with GenBank Accession Number
  AC108190 clone RP86-44312 (center project name cod)"
```

##### misc\_feature

```
58366..138683
  /note="clone overlaps with GenBank Accession Number
  AC108899 clone RP86-261015 (center project name cof)"
```

##### misc\_feature

```
105553..138683
  /note="assembly_fragment"
  /clone_end:F7
  /vector_side:right"
```

#### ORIGIN

```
Query Match      16.9%; Score 40.2; DB 2; Length 138683;
Best Local Similarity 52.0%; Pred. No. 2.4;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 55 TTGGCACTGCTATTATTCATTCGCAATCATTTCGGATGTTCTCGAAGAGCTTC 114
DB 41781 TTTGCTTCTCCAGATACACTTCTCTTACAGTTATGGATAGTCACACCTATTTC 41722

QY 115 CCBAAGTTATTGGAGTACTGTGMAAGAGTTTCGTCATGAAGTTTACCCAAAGGACTTTACT 174
DB 41721 ACACACTGTGAGAGTTTAAATGAGTAAATGATGATGATTAATGATTAATGATTAATGATGATGAT 41662

QY 175 ATGTGAATTAATTTGTCAAACACTAGTAGTCAGATCAATAAATTTCTACSTGGCA 227
DB 41661 AGTTAAACAGTAATAGTAAACAGTAGTAATAATAATAATAATAATAATTTGCTTAGGCA 41609
```

#### RESULT 8

##### LOCUS

AC126197/c  
Rattus norvegicus clone CH230-145B6, WORKING DRAFT SEQUENCE, 3  
unordered pieces.

##### ACCESSION

AC126197 GI:30579275

##### KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.



## SOURCE Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

## AUTHORS

2 (bases 1 to 250470)  
Xuzhu, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayayi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, X., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deranc, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Fieley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Georgiadis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, X.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Herrandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenstunewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Xiner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, D.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwankwelenh, O., Okwono, G., Olarnpunagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plommer, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steimle, X., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,  
Valas, R., Vera, V., Villalana, J., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willison, R., Wleczyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

2 (bases 1 to 250470)  
Worley, K.C.  
Direct Submission  
Submitted (04-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 250470)  
Rat Genome Sequencing Consortium.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:24942206.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GNMK

Center clone name: CH230-145E6

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 240374 bases at least Q40

Consensus quality: 242949 bases at least Q30

Consensus quality: 245059 bases at least Q20

Estimated insert size: 254162; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))

\* NOTE: This sequence may represent more than one clone.

\* NOTE: This is a 'working draft' sequence. It currently

consists of 3 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

\* 1 248046: contig of 248046 bp in length

\* 248047 248146: gap of unknown length

\* 248147 249193: contig of 1047 bp in length

\* 249194 249293: gap of unknown length

\* 249294 250470: contig of 1177 bp in length.

location/Qualifiers

1..250470

FEATURES

source

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-145E6"

misc\_feature

1..2048

/note="wgs contig"

misc\_feature

2721..4418

/note="wgs contig"

misc\_feature

238284..240199

/note="wgs contig"

misc\_feature

241389..245579

/note="wgs contig"

ORIGIN

Query Match 16.5%; Score 39.2; DB 2; Length 250470;

Best Local Similarity 52.4%; Pred. No. 4.2;

Matches 86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 53 AATGCGCACTGTCTATTATTCATTCCGCAATCATTTCGGATGTTCTCGAAGAGACT 112

Db 108876 AATTTATATATAGTATATTTAAAGGTTTCATATTTTAAATGATATATTTGAAATGGACT 108817

QY 113 TCCCAAGCTTATTCGCTACTGTGTAAGAGTTCCTGATGAAGTTTACCAAGAGACTTTA 172

Db 108816 TTCAAAACATATAGGAGAACTGCAAGAAATTTTAAAGGAGATTTATTGATTAATATG 108757

QY 173 CTATGTGATTAATTAATTTGTCAAACTTAGTCTAGTCAGATCAATTAAT 216

Db 108756 AATATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 108713

```

RESULT 9
AC146595/c
LOCUS AC146595.1 177864 bp DNA linear HTG 06-SEP-2003
DEFINITION Mus musculus chromosome UNK clone RP24-164G18, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AC146595
VERSION AC146595.1 GI:3494968
KEYWORDS HTG: HTGS PHASE1.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 177864)
AUTHORS Wilson,R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177864)
AUTHORS Wilscot,R.K.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information
Center project name: M_BB01543:8
----- Summary Statistics
Sequencing vector: M13; 0%
Sequencing method: plasmid; 100%
Chemistry: Dye-terminator ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175447 bases at least Q40
Consensus quality: 175803 bases at least Q33
Consensus quality: 175912 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1486: contig of 1486 bp in length
* 1487 1586: gap of unknown length
* 1587 24281: contig of 22695 bp in length
* 24282 24381: gap of unknown length
* 24382 60849: contig of 36468 bp in length
* 60850 60949: gap of unknown length
* 60950 175073: contig of 114124 bp in length
* 175074 175173: gap of unknown length
* 175174 176585: contig of 1412 bp in length
* 176586 176685: gap of unknown length
* 176686 177864: contig of 1179 bp in length.
----- Location/Qualifiers
1. 177864
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-164G18"
-----
misc_feature 1. 1486
/note="assembly_name:Contig12"
misc_feature 1587. 24281
/note="assembly_name:Contig13"
misc_feature 24382. 60849
/note="assembly_name:Contig14"
misc_feature 60950. 175073
-----
/note="assembly_name:Contig15"
175174. 176585
/note="assembly_name:Contig17"
176686. 177864
/note="assembly_name:Contig19"
-----
misc_feature 16.4%; Score 39; DB 2; Length 177864;
Best Local Similarity 52.8%; Pred. No. 4.9;
Matches 84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Query Match
Db 145250 ATATCGAATCATCTCTAGTGTGTTTGTCAAAATTCACGAGAAAGTAATGAGTTAA 145191
QY 75 ATTGCGAATCATCTCTAGTGTGTTTGTCAAAATTCACGAGAAAGTAATGAGTTAA 134
QY 135 TGAAGAGTTCGTCATGAGTTTACCCAAAGGACTTACTATGTGAATTAATTTGCAAA 194
Db 145190 TTCACTTTTATTACTATTAATTCGAAATAATTTAAATTTATATATTAACATCAAA 145131
QY 195 CTAGTAGTCAGATCAATTAATTTCTAGTGTGCAAAATA 233
Db 145130 CTCATAACAATCTCTTTTGTGTTTATTATTGCGCAATAAA 145092
-----
RESULT 10
AC108074
LOCUS AC108074 128635 bp DNA linear PRI 20-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-717H13 from 4, complete sequence.
ACCESSION AC108074 AC024668
VERSION AC108074.3 GI:19551225
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 128635)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 128635)
AUTHORS Trani,L., Cotton,M., Creason,K. and Nguyen,C.
TITLE The sequence of Homo sapiens BAC clone RP11-717H13
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 128635)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 128635)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 128635)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
On Mar 20, 2002 this sequence version replaced gi:16702446.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0717H13
Drafting Center: WMBR
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanesse, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-384D10, 2000 bp overlap; the clone sequenced to the right is RP11-766L18. Actual start of this clone is at base position 21951 of RP11-384D10; actual end is at base position 128635 of RP11-717H13.

Data from AC027411 was used to finish this clone, AC108074.

The sequence of AC024668 has been incorporated into AC108074.

#### FEATURES

source

1. 128635  
 /location="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-717H13"  
 /clone\_lib="RPCI-11"  
 /rpt\_family="Alu"  
 7. 317  
 /rpt\_family="Alu"  
 400. 449  
 /rpt\_family="L2"  
 746. 811  
 /rpt\_family="(TA)n"  
 829. 1006  
 /rpt\_family="(TA)n"  
 1007. 1030  
 /rpt\_family="(CA)n"  
 1275. 1708  
 /rpt\_family="MaLR"  
 2046. 2066  
 /rpt\_family="(Alu)n"  
 3093. 3133  
 /rpt\_family="AT\_rich"  
 3735. 4324  
 /rpt\_family="L1"  
 483. 5375  
 /rpt\_family="ERV1"  
 5422. 6079  
 /rpt\_family="ERV1"  
 6486. 6619  
 /rpt\_family="ERV1"  
 8484. 8649  
 /rpt\_family="L2"

repeat\_region 8682. 9151  
 /rpt\_family="L1"  
 repeat\_region 9152. 10608  
 /rpt\_family="L1"  
 repeat\_region 10658. 10904  
 /rpt\_family="L2"  
 repeat\_region 11911. 12211  
 /rpt\_family="Alu"  
 repeat\_region 13330. 13635  
 /rpt\_family="Alu"  
 repeat\_region 14025. 14067  
 /rpt\_family="AT\_rich"  
 repeat\_region 14979. 15008  
 /rpt\_family="(TG)n"  
 repeat\_region 15622. 15927  
 /rpt\_family="MaLR"  
 repeat\_region 18551. 19278  
 /rpt\_family="ERV1"  
 repeat\_region 20185. 20238  
 /rpt\_family="AT\_rich"  
 repeat\_region 20443. 20932  
 /rpt\_family="L1"  
 repeat\_region 20911. 21112  
 /rpt\_family="L1"  
 repeat\_region 21135. 23419  
 /rpt\_family="L1"  
 repeat\_region 24585. 24745  
 /rpt\_family="L1"  
 repeat\_region 24746. 25029  
 /rpt\_family="Alu"  
 repeat\_region 25030. 25916  
 /rpt\_family="L1"  
 repeat\_region 25937. 26213  
 /rpt\_family="L1"  
 repeat\_region 26989. 27013  
 /rpt\_family="AT\_rich"  
 repeat\_region 27206. 27471  
 /rpt\_family="L1"  
 repeat\_region 27609. 27719  
 /rpt\_family="AT\_rich"  
 repeat\_region 30364. 30389  
 /rpt\_family="(TTG)n"  
 repeat\_region 32127. 32803  
 /rpt\_family="L1"  
 repeat\_region 32843. 34316  
 /rpt\_family="L1"  
 repeat\_region 34453. 34952  
 /rpt\_family="L1"  
 repeat\_region 35071. 35154  
 /rpt\_family="ACHobo"  
 repeat\_region 35233. 35457  
 /rpt\_family="L1"  
 repeat\_region 35968. 36083  
 /rpt\_family="L1"  
 repeat\_region 37216. 37395  
 /rpt\_family="MER2\_type"  
 repeat\_region 37398. 37520  
 /rpt\_family="Alu"  
 misc\_feature 37914. 38431  
 /note="match to EST AA809063 (NID:g2878469) nw17a12.s1"  
 repeat\_region 39777. 41891  
 /rpt\_family="L1"  
 repeat\_region 41892. 42231  
 /rpt\_family="L1"  
 repeat\_region 42388. 43112  
 /rpt\_family="L1"  
 repeat\_region 43108. 43308  
 /rpt\_family="L1"  
 repeat\_region 43309. 43381  
 /rpt\_family="MER2\_type"  
 repeat\_region 43382. 45504  
 /rpt\_family="L1"  
 repeat\_region 46562. 46650

```

/rpt_family="CP-rich"
46753..46938
/rpt_family="(ATATG)n"
46884..47008
/rpt_family="(TA)n"
48559..48780
/rpt_family="Li"

```

```

Query Match      16.3%; Score 38.8; DB 9; Length 128635;
Best Local Similarity 55.1%; Pred. No. 5.7;
Matches %; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 81 CAAATCATTTCGAGTCTTCGAAAGGACTTCCAAAGTTATTGAGTACTGTTGAAAG 140
    |||||
Db 110386 CAATGAATTTCAACATATTCGAGAGAGGATTCAGAAATTCATCAGAGAAATTAACAG 110445
    |||||

QY 141 AGTTCGATCAGATTCACCAAGGACTTTACTATGTAATTAATTCACACTACTA 200
    |||||
Db 110446 AGATTAATTAATTTTAAATCAAGCAGGAAATCTGGAGCCAAATGTTCAACTGGGA 110505
    |||||

QY 201 GTCAGATCAATAAAATTC 218
    |||||
Db 110506 AAATGAATAAATAATAC 110523
    |||||

```

```

RESULT 11
AP004700/c
LOCUS      138772 bp DNA linear HTG 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 8 clone P0518H02,
            *** SEQUENCING IN PROGRESS ***.
ACCESSION  AP004700
VERSION     AP004700.1 GI:18447959
KEYWORDS   HTG; HTGS PHASE2.
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM   Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.

```

```

REFERENCE  1. Sasaki, T., Matsumoto, T. and Yamamoto, K.
            Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, PAC
            clone: P0518H02
JOURNAL    Published Only in Database (2002)
REFERENCE  2. (bases 1 to 138772)
AUTHORS    Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE      Direct Submission
JOURNAL    Submitted (30-JAN-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program, Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            [E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
            Tel: 81-298-38-7441, Fax: 81-298-38-7468]
COMMENT    NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.

```

```

FEATURES
            Location/Qualifiers
            1..138772
               /organism="Oryza sativa (japonica cultivar-group)"
               /mol_type="genomic DNA"
               /cultivar="Nipponbare"
               /db_xref="taxon:39947"
               /chromosome="8"
               /clone="P0518H02"

```

## ORIGIN

```

Query Match      15.3%; Score 38.8; DB 2; Length 138772;
Best Local Similarity 55.4%; Pred. No. 5.6;

```

```

Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 133 TGTCAAGAGAGTTGTCATGAAGTTTACCAAGAGACTTTACTATGTAATTAATTTGCA 192
    |||||
Db 70260 TGTCAACATTTTGAACGGGGAATATAAGGATTTTAACTTTTACTTACTACTGTT 70201
    |||||

QY 133 AACTAGTAGTCAGATCAATCAATTTCTACGTGGCAAAAAA 238
    |||||
Db 70200 AACCAATTTGTTTATTAATAAAATTTATACAAATATAAAAAATAAAA 70155
    |||||

```

```

RESULT 12
AP005757/c
LOCUS      147655 bp DNA linear PLN 26-SEP-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
            PAC clone: P0413H11, complete sequence.
ACCESSION  AP005757
VERSION     AP005757.2 GI:35764583
KEYWORDS   HTG.
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM   Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.

```

```

REFERENCE  1. Sasaki, T., Matsumoto, T. and Katayose, Y.
            Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, PAC
            clone: P0413H11
JOURNAL    Published Only in Database (2002)
REFERENCE  2. (bases 1 to 147655)
AUTHORS    Sasaki, T., Matsumoto, T. and Katayose, Y.
TITLE      Direct Submission
JOURNAL    Submitted (18-SEP-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program, Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            [E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
            Tel: 81-298-38-7441, Fax: 81-298-38-7468]

```

```

COMMENT    On Sep 25, 2003 this sequence version replaced gi:23200616.
            The orientation of the sequence is from 17 to SP6 of the PAC clone.
            Location/Qualifiers
            1..147655
               /organism="Oryza sativa (japonica cultivar-group)"
               /mol_type="genomic DNA"
               /cultivar="Nipponbare"
               /db_xref="taxon:39947"
               /chromosome="8"
               /clone="P0413H11"

```

## FEATURES

```

            Location/Qualifiers
            1..147655
               /organism="Oryza sativa (japonica cultivar-group)"
               /mol_type="genomic DNA"
               /cultivar="Nipponbare"
               /db_xref="taxon:39947"
               /chromosome="8"
               /clone="P0413H11"

```

## ORIGIN

```

Query Match      16.3%; Score 38.8; DB 8; Length 147655;
Best Local Similarity 60.4%; Pred. No. 5.6;
Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 133 TGTCAAGAGAGTTGTCATGAAGTTTACCAAGAGACTTTACTATGTAATTAATTTGCA 192
    |||||
Db 128424 TGTCAACATTTTGAACGGGGAATATAAGGATTTTAACTTTTACTTACTACTGTT 128365
    |||||

QY 193 AACTAGTAGTCAGATCAATCAATTTCTACGTGGCAAAAAA 238
    |||||
Db 128364 AACCAATTTGTTTATTAATAAAATTTATACAAATATAAAAAATAAAA 128319
    |||||

```

```

RESULT 13
AC132361/c
LOCUS      161010 bp DNA linear HTG 33-SEP-2002
DEFINITION Mus musculus chromosome UNK clone RP24-149E13, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION  AC132361
VERSION     AC132361.1 GI:22657848
KEYWORDS   HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

```



/db xref="taxon:95105"  
 <1..>1162  
 /product="28S ribosomal RNA"

## rRNA

## ORIGIN

Query Match 16.2%; Score 38.6; DB 3; Length 1182;  
 Best Local Similarity 51.4%; Pred. No. 9.7;  
 Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
 QY 34 GATGCAAGCTCTCTCAATGCGCAACTGCTATTATTCATTCGCCAATCACAATTCG 93  
 DB 224 GATGCAAGCTCTCTCAATGCGCAACTGCTATTATTCATTCGCCAATCACAATTCG 283  
 QY 94 GATGTTCTCGAAAGGACTTCCCAAGTATTGGAGTACTGTGAAGAGTGTGTCAAGAA 153  
 DB 284 GAAATTTATAGAAATATTTTATATAGTTTATATTAAGTAATTAAGTTGGTATGAC 343  
 QY 154 GTTTCACCAAGACTTACTATGTGATTAATTTCTCAACTAGTAGTACAG 206  
 DB 344 TCTTCACATGAGTATGATCCAAATTCAGGACATTCGCCAGTTGGAGTTTGA 396

## RESULT 15

AC108190/c 132389 bp DNA linear HTG 19-NOV-2002  
 LOCUS Felis catus clone RP86-44312, WORKING DRAFT SEQUENCE.  
 DEFINITION AC108190  
 ACCESSION AC108190.2 GI:25101059  
 VERSION HTG: HTGS PHASE2; HTGS\_DRAFT.  
 KEYWORDS Felis catus (cat)  
 SOURCE Felis catus  
 ORGANISM Felis catus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 1. (bases 1 to 132389)  
 Akhter N., Antonellis A., Ayele K., Beckstrom-Sternberg S.M.,  
 Benjamin B., Blakesley R.W., Bouffard G.G., Brinkley C., Brooks S.,  
 Carliaga K., Coleman B., Engle J., Granite S., Guan X., Gupta J.,  
 Haghighi P., Han J., Hansen N., Ho S.-L., Idol J.R., Karlins E.,  
 Laric P., Lee-Lin S.-Q., Legaspi R., Maduro Q.B., Maduro V.B.,  
 Margulies E.H., Masiello C., Maskeri B., McDowell J.,  
 Padirigan C., Pearson R., Portney M.E., Prasad A.,  
 Reddix-Dugue N., Schandler K., Schuefer M.G., Sison C.,  
 Stancipop S., Thomas J.W., Thomas P., Touchman J.W., Vogt J.L.,  
 Wetherby K.D., Wiggins L., Young A. and Green E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2. (bases 1 to 132389)  
 Green E.D.

## AUTHORS

Direct Submission  
 Submitted (26-JAN-2002) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 3. (bases 1 to 132389)  
 Green E.D.  
 Direct Submission  
 Submitted (19-NOV-2002) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 On Nov 19, 2002 this sequence version replaced gi:18376889.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nsc.zoo@ngri.nih.gov](mailto:nsc.zoo@ngri.nih.gov)  
 ----- Project Information  
 Center project name: cod  
 Center clone name: 443102

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

## ----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 132361 bases at least Q40  
 Consensus quality: 132381 bases at least Q30  
 Consensus quality: 132389 bases at least Q20  
 Insert size: 134000; agarose-fp  
 Insert size: 132389; sum-of-contigs  
 Quality coverage: 11.10x in Q20 bases; agarose-fp  
 Quality coverage: 11.23x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* I 132389: contig of 132389 bp in length.

## FEATURES

## source

1. 132389  
 /organism="Felis catus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9685"  
 /clone="RP86-44312"  
 /clone\_lib="RP86"

## misc\_feature

1. 132389  
 /note="assembly\_fragment  
 clone end:17  
 vector side:left  
 clone end:Sp6  
 vector side:right"

## misc\_feature

1. 70573  
 /note="clone overlaps with GenBank Accession Number  
 AC108196 clone RP86-591N22 (center project name cod)"  
 70580..132389

## misc\_feature

/note="clone overlaps with GenBank Accession Number  
 AC108193 clone RP86-469M8 (center project name cod)"  
 129445..132389

## misc\_feature

/note="clone overlaps with GenBank Accession Number  
 AC108899 clone RP86-261O15 (center project name cod)"

## ORIGIN

Query Match 16.2%; Score 38.6; DB 2; Length 132389;  
 Best Local Similarity 51.4%; Pred. No. 6.4;  
 Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
 QY 55 TTGGCAACTGCTCTATTATTCATTCGCCAATCACAATTCGGATGTTCTCGAAAGACTTC 114  
 DB 112844 TTGGTCTCTCAGATACACTTCTCTTACAGATTATGGATAGTAACAACACTTATTC 112785  
 QY 115 CCAAGTATTGGAGTACTGTGAAGAGTTCGTCATGAAGTTTACCAAGGACTTACT 174  
 DB 112784 ACACACTGTGAGAGTTAAATGAGGTAAATGCAATAAAGATTAATTAACCTGGTCTTGT 112725  
 QY 175 AATGCAATTAATTTGTCAACTAGTACATCAATTAATTTCTACGTGGCA 227  
 DB 112724 ACTTAACAGTAATAGTACAGTATTAATTAATTAATTAATTTCTAGGCA 112672

Search completed: July 7, 2004, 15:39:07  
 Job time : 1868 secs

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition,

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 7, 2004, 13:19:03 ; Search time 2509 Seconds

(without alignments)  
2832.682 Million cell updates/sec

Title: US-09-437-450A-40

Perfect score: 238

Sequence: 1 tttttttttttggagga.....tagtggcaaaaaaaaaa 238

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026573

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_esthum:\*\*
- 4: em\_esthum:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_hic:\*\*
- 9: gb\_est1:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_hic:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: em\_gss\_hum:\*\*
- 18: em\_gss\_inv:\*\*
- 19: em\_gss\_pln:\*\*
- 20: em\_gss\_vrt:\*\*
- 21: em\_gss\_fun:\*\*
- 22: em\_gss\_mam:\*\*
- 23: em\_gss\_mus:\*\*
- 24: em\_gss\_pro:\*\*
- 25: em\_gss\_rod:\*\*
- 26: em\_gss\_ptg:\*\*
- 27: em\_gss\_vrt:\*\*
- 28: gb\_gss1:\*\*
- 29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	238	100.0	238	9 AF071356	AF071356 AF071356
2	230.2	96.7	933	9 AF071357	AF071357 AF071357
3	223.2	93.8	238	9 AF071379	AF071379 AF071379
4	160.2	67.3	217	9 AF071398	AF071398 AF071398

5	159	66.8	214	9 AF071362	AF071362 AF071362
6	127	53.4	511	12 BI175173	BI175173 OSTR168D4
7	91	38.2	583	14 CB396206	CB396206 OSTR168D4
8	89.2	37.5	300	9 AUI15788	AUI15788 AUI15788
9	82	34.5	275	14 T01325	T01325 WEST02646 E
10	78.4	32.9	567	14 CB396141	CB396141 CSTR167C8
11	78.4	32.9	567	14 CB396151	CB396151 CSTR167C8
12	77.2	32.4	574	14 CB397715	CB397715 CSTR193H1
13	77.2	32.4	574	14 CB397765	CB397765 CSTR193H1
14	43	18.1	574	29 CE051398	CE051398 tigr-gss-
15	40.8	17.1	636	29 CE717539	CE717539 tigr-gss-
16	40	16.8	334	14 CF245758	CF245758 Fcylcolida
17	39.6	16.6	462	28 AQ862746	AQ862746 ndeb0019C
18	38.6	16.2	663	14 CB443694	CB443694 694795 MA
19	38.6	16.2	695	14 CB444053	CB444053 695179 MA
20	38	16.0	581	14 CF587970	CF587970 USDA-PP-1
21	38	16.0	1201	13 BX353711	BX353711 BX353711
22	37.8	15.9	686	28 AZ512342	AZ512342 LM0357L17
23	37.6	15.8	300	9 AA007439	AA007439 zh98b12.8
24	37.6	15.8	325	28 BZ876771	BZ876771 CH240.278
25	37.6	15.8	455	28 AZ015157	AZ015157 RPCI-23-3
26	37	15.5	358	13 BX569552	BX569552 BX569552
27	37	15.5	632	14 CB431533	CB431533 607523 MA
28	37	15.5	695	14 CB430835	CB430835 606755 MA
29	37	15.5	953	29 CC399587	CC399587 ZUBAH70TV
30	36.6	15.4	354	13 BY302654	BY302654 BY302654
31	36.2	15.2	412	12 BI863409	BI863409 r104f09.Y
32	36.2	15.2	440	13 PQ837948	PQ837948 rfc3h02.Y
33	36	15.1	900	29 CG767772	CG767772 TCB50.2.D
34	35.8	15.0	622	14 CA589048	CA589048 hab53b06
35	35.6	15.0	599	14 CF440904	CF440904 EST677249
36	35.4	14.9	332	14 CB393253	CB393253 OSTR116A3
37	35.4	14.9	590	28 BH186791	BH186791 C32_K_23-
38	35.4	14.9	590	29 CNS07R3V	CNS07R3V T3 end of
39	35.4	14.9	1069	13 BQ899371	BQ899371 AGENCOURT
40	35.2	14.8	427	13 BQ602354	BQ602354 MI-P-HO-a
41	35.2	14.8	432	12 BMS37674	BMS37674 hab5c01.g
42	35.2	14.8	434	9 AA532321	AA532321 QEST.451
43	35.2	14.8	492	28 BH312146	BH312146 CH230-101
44	35.2	14.8	681	13 BX676706	BX676706 BX676706
45	35.2	14.8	726	28 BH979578	BH979578 cdf92908.

ALIGNMENTS

RESULT 1	AF071356	AF071356	238 bp	mRNA	linear	EST 25-NOV-1998
LOCUS	AF071356	AF071356	mRNA from cadmium-responsive gene	Caenorhabditis elegans		
DEFINITION	AF071356	AF071356	cDNA clone DORT16, mRNA sequence.			
ACCESSION	AF071356	AF071356.1	GI:3265101			
VERSION	AF071356	AF071356.1	GI:3265101			
KEYWORDS	AF071356	AF071356.1	GI:3265101			
SOURCE	AF071356	AF071356.1	GI:3265101			
ORGANISM	AF071356	AF071356.1	GI:3265101			
REFERENCE	AF071356	AF071356.1	GI:3265101			
AUTHORS	AF071356	AF071356.1	GI:3265101			
TITLE	AF071356	AF071356.1	GI:3265101			
JOURNAL	AF071356	AF071356.1	GI:3265101			
MEDLINE	AF071356	AF071356.1	GI:3265101			
PUBMED	AF071356	AF071356.1	GI:3265101			
COMMENT	AF071356	AF071356.1	GI:3265101			
1	(bases 1 to 238)					
Liao, V.H.-C. and Freedman, J.H.						
Cadmium-regulated genes from the nematode <i>Caenorhabditis elegans</i> .						
Identification and cloning of new cadmium-responsive genes by						
differential display						
J. Biol. Chem. 273 (48), 31962-31970 (1998)						
Contact: Jonathan H. Freedman						
Nicholas School of the Environment						
Duke University						
Box 90328, Durham, NC 27708-0328, USA						
Email: jon@duke.edu						
Location/Qualifiers						
I. .238						

FEATURES  
source

```
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="DDRT16"
/tissue_type="whole animal"
/dev_stage="mixed population"
/clone_lib="mRNA from cadmium-responsive gene"

Query Match 100.0%; Score 238; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.5e-45;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTCTTTTGGGAGGAATCAGCGCGGATGCAACAGTCTTCTCAATGGCACT 758
Db 1 TTTTCTTTTGGGAGGAATCAGCGCGGATGCAACAGTCTTCTCAATGGCACT 758
64 GTCTATTATCCATTCGCAATCACATTTTCGGATGTTCTCGAAAGGACTTCCCAAGTTA 123
Db 64 GTCTATTATCCATTCGCAATCACATTTTCGGATGTTCTCGAAAGGACTTCCCAAGTTA 123
759 GTCTATTATCCATTCGCAATCACATTTTCGGATGTTCTCGAAAGGACTTCCCAAGTTA 818
Db 759 GTCTATTATCCATTCGCAATCACATTTTCGGATGTTCTCGAAAGGACTTCCCAAGTTA 818
124 TTGGAGTACTGTGAAGAGTGTGTCATCAAGTTTACCAAGGACTTTTACTATGTAAT 183
Db 124 TTGGAGTACTGTGAAGAGTGTGTCATCAAGTTTACCAAGGACTTTTACTATGTAAT 183
819 TTGGAGTACTGTGAAGAGTGTGTCATCAAGTTTACCAAGGACTTTTACTATGTAAT 878
Db 819 TTGGAGTACTGTGAAGAGTGTGTCATCAAGTTTACCAAGGACTTTTACTATGTAAT 878
184 AATTTCTCAACTAGTAGTCAGATCAATAAATAATTTCTACGTGGCAAAAAA 238
Db 184 AATTTCTCAACTAGTAGTCAGATCAATAAATAATTTCTACGTGGCAAAAAA 238
879 AATTTCTCAACTAGTAGTCAGATCAATAAATAATTTCTACGTGGCAAAAAA 933
Db 879 AATTTCTCAACTAGTAGTCAGATCAATAAATAATTTCTACGTGGCAAAAAA 933

RESULT 3
LOCUS AF071379 238 bp mRNA linear EST 25-NOV-1998
DEFINITION AF071379 mRNA from cadmium-responsive gene Caenorhabditis elegans
cDNA clone DRT16, mRNA sequence.
ACCESSION AF071379
VERSION AF071379.1 GI:3265124
KEYWORDS EST
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 238)
AUTHORS Liao, V.H.-C. and Freedman, J.H.
TITLE Cadmium-regulated genes from the nematode Caenorhabditis elegans.
Identification and cloning of new cadmium-responsive genes by
differential display
J. Biol. Chem. 273 (48), 31962-31970 (1998)
JOURNAL 99041962
MEDLINE 9822667
PUBMED
COMMENT Contact: Jonathan H. Freedman
Nicholas School of the Environment
Duke University
Box 90328, Durham, NC 27708-0328, USA
Email: jhf@duke.edu.

FEATURES
source
1..238
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/db_xref="taxon:6239"
/clone="DDRT16"
/tissue_type="whole animal"
/dev_stage="mixed population"
/clone_lib="mRNA from cadmium-responsive gene"

ORIGIN
Query Match 93.8%; Score 223.2; DB 9; Length 238;
Best Local Similarity 96.6%; Pred. No. 4.2e-42;
Matches 228; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TTTTCTTTTGGGAGGAATCAGCGCGGATGCAACAGTCTTCTCAATGGCA 60
Db 1 TTTTCTTTTGGGAGGAATCAGCGCGGATGCAACAGTCTTCTCAATGGCA 60
61 ACTGTCATTATCCATTCGCAATCACATTTTCGGATGTTCTCGAAAGGACTTCCCAAG 120
Db 61 ACTGTCATTATCCATTCGCAATCACATTTTCGGATGTTCTCGAAAGGACTTCCCAAG 120
121 TTATGGAGTACTGTGAAGAGTGTGTCATCAAGTTTACCAAGGACTTTACTATGTA 180
Db 121 TTATGGAGTACTGTGAAGAGTGTGTCATCAAGTTTACCAAGGACTTTACTATGTA 180
181 ATTAATTTGCAACTAGTAGTCAGATCAATAAATAATTTCTACGTGGCAAAAAA 238
Db 181 ATTAATTTGCAACTAGTAGTCAGATCAATAAATAATTTCTACGTGGCAAAAAA 238

RESULT 2
LOCUS AF071357 933 bp mRNA linear EST 25-NOV-1998
DEFINITION AF071357 mRNA from cadmium-responsive gene Caenorhabditis elegans
cDNA clone DRT16FC, mRNA sequence.
ACCESSION AF071357
VERSION AF071357.1 GI:3265102
KEYWORDS EST
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 933)
AUTHORS Liao, V.H.-C. and Freedman, J.H.
TITLE Cadmium-regulated genes from the nematode Caenorhabditis elegans.
Identification and cloning of new cadmium-responsive genes by
differential display
J. Biol. Chem. 273 (48), 31962-31970 (1998)
JOURNAL 99041962
MEDLINE 9822667
PUBMED
COMMENT Contact: Jonathan H. Freedman
Nicholas School of the Environment
Duke University
Box 90328, Durham, NC 27708-0328, USA
Email: jhf@duke.edu.

FEATURES
source
1..933
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/db_xref="taxon:6239"
/clone="DRT16FC"
/tissue_type="whole animal"
/dev_stage="mixed population"
/clone_lib="mRNA from cadmium-responsive gene"

ORIGIN
Query Match 96.7%; Score 230.2; DB 9; Length 933;
Best Local Similarity 98.7%; Pred. No. 5e-44;
Matches 232; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TTTTCTTTTGGGAGGAATCAGCGCGGATGCAACAGTCTTCTCAATGGCACT 53
Db 4 TTTTCTTTTGGGAGGAATCAGCGCGGATGCAACAGTCTTCTCAATGGCACT 53
```



```

RESULT 4
AF071398      217 bp  mRNA  linear  EST 25-NOV-1998
LOCUS
DEFINITION AF071398 mRNA from cadmium-responsive gene Caenorhabditis elegans
cDNA clone DDX17, mRNA sequence.
ACCESSION AF071398
VERSION AF071398
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 217)
AUTHORS Liao V.H.-C. and Freedman, J.H.
TITLE Cadmium-regulated genes from the nematode Caenorhabditis elegans.
Identification and cloning of new cadmium-responsive genes by
differential display
J. Biol. Chem. 273 (48), 31962-31970 (1998)
JOURNAL 99041962
MEDLINE 9822667
PUBMED
COMMENT Contact: Jonathan H. Freedman
Nicholas School of the Environment
Duke University
Box 90328, Durham, NC 27708-0328, USA
Email: jonf@duke.edu.

FEATURES
source
1..217
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="DDRT2"
/tissue_type="whole animal"
/dev_stage="mixed population"
/clone_lib="mRNA from cadmium-responsive gene"

ORIGIN
Query Match 67.3%; Score 160.2; DB 9; Length 217;
Best Local Similarity 92.2%; Pred. No. 2.4e-27;
Matches 202; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

Qy 1 TTTTCTTTTGGAGGAAATCAGCGCGGATGCAACAGTCTCTCTCAATTGGCA 60
Db 2 TTTTCTTTTGGAGGAAATCAGCGCGGATGCAACAGTCTCTCTCAATTGGCA 61

Qy 61 ACTGCTATTATCCATTCGCAATTCAGTTCGGATGTTCTCGAAGGACTTCCAAAG 120
Db 62 ACTGCTAT--ATCATTCGCAATTCAGTTCGGATGTTCTCGAAGG--CATTCCAAG 118

Qy 121 TTATTGAGTACTGTGAAGAGTTCTGTCATGAAGTTTACCCAAAGGACTTACTATGTA 180
Db 119 TTATTGAGTACTGTGAAGAGTTCTGTCATGAAGTTTACCCAAAGGACTTACTATGTA 177

Qy 181 ATTAATGTCAACTAGTACTGAGTCAATTAATTTCT 219
Db 178 ATTAATGTCAACTAGTACTGAGTCAATTAATTTCT 216

RESULT 5
AF071362      214 bp  mRNA  linear  EST 25-NOV-1998
LOCUS
DEFINITION AF071362 mRNA from cadmium-responsive gene Caenorhabditis elegans
cDNA clone DDX2, mRNA sequence.
ACCESSION AF071362
VERSION AF071362
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 214)
AUTHORS Liao, V.H.-C. and Freedman, J.H.

```

```

TITLE Cadmium-regulated genes from the nematode Caenorhabditis elegans.
Identification and cloning of new cadmium-responsive genes by
differential display
J. Biol. Chem. 273 (48), 31962-31970 (1998)
JOURNAL 99041962
MEDLINE 9822667
PUBMED
COMMENT Contact: Jonathan H. Freedman
Nicholas School of the Environment
Duke University
Box 90328, Durham, NC 27708-0328, USA
Email: jonf@duke.edu.

FEATURES
source
1..214
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="DDRT2"
/tissue_type="whole animal"
/dev_stage="mixed population"
/clone_lib="mRNA from cadmium-responsive gene"

ORIGIN
Query Match 66.8%; Score 159; DB 9; Length 214;
Best Local Similarity 95.0%; Pred. No. 4.6e-27;
Matches 208; Conservative 0; Mismatches 5; Indels 6; Gaps 4;

Qy 20 AATCAACGCGCGGATGCAACAGTCTCTCTCAATGCAACTCTCTTATTATCAATCC 79
Db 1 AATCATGCGCGGATGCAACAGTCT--TCTCAATTTGGCAACTCTCTATAT--CATTC 56

Qy 80 GCAATCAATTCGGATGTTCTCGAAGAGGACTTCCCAAGTATTTGAGTACTGTGAAA 139
Db 57 GCAA--CACATTTGCGATGTTCTCGAAGAGGACTTCCCAAGTATTTGAGTACTGTGAAA 115

Qy 140 GAGTTCGTCATGAAGTTTACCCAAAGGACTTACTATGTAATTAATTTGCAAACTAGT 199
Db 116 GAGTTCGTCATGAAGTTT--CCCAAGGACTTACTATGTAATTAATTTGCAAACTAGT 174

Qy 200 AGTCAGATCAATAAATTTCTAGTCGCAAAAAA 238
Db 175 AGTCAGATCAATAAATTTCTAGTCGCAAAAAA 213

RESULT 6
B1175173/c 511 bp  mRNA  linear  EST 09-JUL-2001
LOCUS
DEFINITION B1175173 1 AD-wm:DNA Caenorhabditis elegans cDNA similar to
F35E8.11, mRNA sequence.
ACCESSION B1175173
VERSION B1175173.1 GI:14640976
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 511)
AUTHORS Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T.,
Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J.,
Lee, H., Hitti, J., Doucette-Stamm, L., Hartley, J. L., Temple, G. F.,
Brasch, M. A., Vandenhaute, J., Lamesch, P. E., Hill, D. E. and Vidal, M.
Open-reading-frame sequence tags (OSTs) support the existence of at
least 17,300 genes in C. elegans
Nat. Genet. 27 (3), 332-336 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT Contact: Reboul, J, Vaglio P
Marc Vidal Laboratory
Dana Farber Cancer Institute
44 Binney Street, Boston, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 2425
Email: Jerome_Reboul@dfci.harvard.edu

```

Sequence tag of Gateway entry clones. The primers used were designed on the predicted protein encoding ORF. C. elegans ORFome cloning project : Contact jerome\_reboul@fci.harvard.edu or philippe\_vaglio@fci.harvard.edu  
POLYA=No.

#### FEATURES

source  
Location/Qualifiers  
1..511  
/organism="Caenorhabditis elegans"  
/mol\_type="mRNA"  
/strain="N2"  
/db\_xref="taxon:6239"  
/sex="Hermaphrodite and male"  
/tissue\_type="whole animal"  
/dev\_stage="mixed stage"  
/clone\_lib="AD-wrmcDNA"  
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

#### ORIGIN

Query Match 53.4%; Score 127; DB 12; Length 511;  
Best Local Similarity 90.1%; Pred. No. 9.1e-20;  
Matches 136; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 4 TTTTCTTTTGGAGGAAATCACGGCGGGGATGCAACAGTCTTCTCTCAATGGCAACT 63  
Db 151 TTTCTTTTGGAGGAAATCACGGCGGGGATGCAACAGTCTTCTCTCAATGGCAACT 92  
QY 64 GTCTATATCCATTCGCATTCGATGTTCTGCAAGAGTCTCCCAAGTTA 123  
Db 91 GTCTATATCCATTCGCATTCGATGTTCTGCAAGAGTCTCCCAAGTTA 32  
QY 124 TTGGAGTACTGCAAGAGTTCGTCATGAG 154  
Db 31 TTGTAGTACTGCAAGAGTTCGTCATGAG 1

#### RESULT 7

CB396206/c  
LOCUS CB396206 583 bp mRNA linear EST 15-MAY-2003  
DEFINITION OSTR168B4\_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.  
ACCESSION CB396206  
VERSION CB396206.1 GI:30737917  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans  
ORGANISM Caenorhabditis elegans

#### REFERENCE

AUTHORS Reboul,J., Vaglio,P., Rual,J.F., Lamesch,P., Martinez,M., Armstrong,C.M., Li,S., Jacotot,L., Bertin,K., Janky,R., Moore,T., Hudson,J.R., Hartley,J.L., Brasch,M.A., Vandenhaute,J., Boulton,S., Endress,S.A., Jenna,S., Chevet,E., Papasotiropoulos,V., Tolias,P., Ptacek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H., Doucette-Stamm,L., Hill,D.E. and Vidal.M.  
C. elegans ORFome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression  
Nac. Genet. (2003) In press  
Contact: Vidal M  
Marc Vidal Laboratory  
Dana Farber Cancer Institute  
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA  
Tel: 617 632 5180  
Fax: 617 632 5739  
Email: Marc.Vidal@fci.harvard.edu

#### TITLE

Sequence tag of Gateway entry clones. The primers used were designed on the predicted protein encoding ORF. C. elegans ORFome cloning project : Contact david\_hill@fci.harvard.edu or marc\_vidal@fci.harvard.edu  
POLYA=No.

#### FEATURES

source  
Location/Qualifiers  
1..583  
/organism="Caenorhabditis elegans"  
/mol\_type="mRNA"  
/strain="N2"  
/db\_xref="taxon:6239"  
/sex="Hermaphrodite and male"  
/tissue\_type="whole animal"  
/dev\_stage="mixed stage"  
/clone\_lib="AD-wrmcDNA"  
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

#### ORIGIN

Query Match 38.2%; Score 91; DB 14; Length 583;  
Best Local Similarity 70.8%; Pred. No. 2.2e-11;  
Matches 121; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 4 TTTTCTTTTGGAGGAAATCACGGCGGGGATGCAACAGTCTTCTCTCAATGGCAACT 63  
Db 176 TATTATTATGGAGACCAATTCCTCCAGTTGATGGACGGTTTTCAGCCAGTTCGAGT 117  
QY 64 GTCTATATCCATTCGCATTCGATGTTCTGCAAGAGTCTCCCAAGTTA 123  
Db 116 GTTATTATTCATTTTACACTCACATCTCTACTGTCTTGTGAAATGATTTCCCAAGAT 57  
QY 124 TTGGAGTACTGCAAGAGTTCGTCATGAGTTCGCAAGAGTCTTCTTACT 174  
Db 56 CTACAGTACTGCAAGAGTTCGCAAGAGTTCGCAAGAGTTCGCAAGTTCATTTACT 6

#### RESULT 8

AU115788/c  
LOCUS AU115788 300 bp mRNA linear EST 19-OCT-2000  
DEFINITION AU115788 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone YK73695 3', mRNA sequence.  
ACCESSION AU115788  
VERSION AU115788.1 GI:10929355  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans  
ORGANISM Caenorhabditis elegans

#### REFERENCE

AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S.  
A complementary view of the C. elegans genome  
Unpublished (2000)  
Contact: Yuji Kohara  
Genome Biology Lab.  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.

#### TITLE

Location/Qualifiers  
1..300  
/organism="Caenorhabditis elegans"  
/mol\_type="mRNA"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="YK73695"  
/sex="Hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="varied"  
/clone\_lib="unpublished oligo-capped cDNA library"

#### ORIGIN

Query Match 37.5%; Score 89.2; DB 9; Length 300;  
Best Local Similarity 70.6%; Pred. No. 8.1e-11;



```

/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

```

ORIGIN

```

Query Match      17.1%; Score 40.8; DB 29; Length 636;
Best Local Similarity 48.3%; Pred. No. 12;
Matches 114; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 3 TTTTCTTTTGGGAGGAAATCAGCGCGGGATGCCAGCTCTCTCTCAATTGGCAAC 62
Db 236 TTTTCTTTTGGGAGGAAATCAGCGCGGGATGCCAGCTCTCTCTCAATTGGCAAC 62
QY 63 TGTCTATTATCCATTCGGCAATCACATTCGGATGTTCTCGAAAAGGACTTCCAAAGTT 122
Db 176 TAGATATTATTATGTAAGATGAGCTTTTCTTTTAAACAATATCCCTTCBAAT 117
QY 123 ATTGGAGTACTGTGAAAGAGTTGGTATGAGTTTACCCAAAGGACTTTACTATGTGAAT 182
Db 116 TTCTAACTACTTTTGAATGCTGATGACTTAACCTGGTGATATAAAGGAGTATTAAAGTCT 57
QY 183 TAAATTGTCAACTAGTACTAGTCAATCAATTAATTTCTAGTGGCAGAAAAGAAA 238
Db 56 ACCTTTCCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1

```

Search completed: July 7, 2004, 16:21:09  
Job time : 2513 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 7, 2004, 11:39:02 ; Search time 349 Seconds  
(without alignments)

2897.052 Million cell updates/sec

Title: US-09-437-450A-40

Perfect score: 238

Sequence: 1 tttttttttttggaggga.....tagctggcagaaaaa 238

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124039041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04: \*

- 1: Geneseqn1980s: \*
- 2: Geneseqn1990s: \*
- 3: Geneseqn2000s: \*
- 4: Geneseqn2001as: \*
- 5: Geneseqn2001bs: \*
- 6: Geneseqn2002s: \*
- 7: Geneseqn2003as: \*
- 8: Geneseqn2003bs: \*
- 9: Geneseqn2003cs: \*
- 10: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.6	14.5	199	5	ABV55254
2	34.6	14.5	6036	6	ABN80125
3	33.6	14.1	671	3	AAZ27331
4	33.4	14.0	404	8	ACH29861
5	33.4	14.0	1208	4	AAD16067
6	33.2	13.9	8781	6	ABL33686
7	33	13.9	20693	4	ABL16740
8	32.8	13.8	571	6	ABX09781
9	32.8	13.8	678	6	ABQ85781
10	32.8	13.8	745	4	AAK87745
11	32.8	13.8	745	4	AAI57549
12	32.8	13.8	745	6	ABS99726
13	32.8	13.8	745	9	ADB92806
14	32.8	13.8	2630	4	ABL29872
15	32.8	13.8	3423	6	ABQ70819
16	32.8	13.8	3581	4	ABL02096
17	32.8	13.8	6759	4	ABL29758
18	32.8	13.8	83391	6	ABQ67094
19	32.6	13.7	2237	6	ABL58167
20	32.6	13.7	6117	6	ABL3449C
21	32.6	13.7	6561	6	ABN80020
22	32.6	13.7	37184	6	ABQ67077
23	32.4	13.6	675	8	AAI57561

24	32.4	13.6	1724	3	AAA76188
25	32.4	13.6	8733	4	ABL29304
26	32.4	13.6	16842	9	ADB54254
27	32.2	13.5	15373	6	ABL32467
28	32.2	13.5	110000	6	ABA90521_04
29	32	13.4	2133	6	ABT07607
30	32	13.4	12987	2	AAI91324
31	31.8	13.4	756	4	AAI96712
32	31.8	13.4	2000	6	ABZ16757
33	31.8	13.4	2443	6	ABJ40935
34	31.8	13.4	6565	4	AAJ46465
35	31.8	13.4	6565	6	ABK31326
36	31.8	13.4	96597	8	ADA03023
37	31.8	13.4	96597	9	ADB72761
38	31.8	13.4	96597	9	ADC85503
39	31.8	13.4	102457	8	ACH03408_3
40	31.6	13.3	264	3	AAA45239
41	31.6	13.3	778	4	AAH33523
42	31.6	13.3	858	4	AAI94027
43	31.6	13.3	4259	3	AAV4870
44	31.6	13.3	5183	2	AAV04063
45	31.6	13.3	6404	6	ABN83979

## ALIGNMENTS

RESULT 1	
ABV55254	
ID	ABV55254 standard; cDNA; 199 BP.
XX	AC
XX	ABV55254;
DT	17-SRP-2002 (first entry)
XX	DE
DE	Human prostate expression marker cDNA 55245.
XX	XX
XX	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX	pharmacogenomic marker; gene; ss.
OS	Homo sapiens.
XX	XX
PN	W200160860-A2.
XX	XX
PD	23-AUG-2001.
XX	XX
PF	20-FEB-2001; 2001WO-US005171.
XX	XX
PR	17-FEB-2000; 2000US-0183319P.
PR	16-MAR-2000; 2000US-0189862P.
PR	25-MAY-2000; 2000US-0207454P.
PR	09-JUN-2000; 2000US-0211314P.
PR	18-JUL-2000; 2000US-0219007P.
PR	13-DEC-2000; 2000US-0255281P.
XX	XX
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	XX
PI	Schlegel R, Endege WO, Monahan JE;
XX	XX
DR	WPI; 2001-662795/76.
XX	XX
PT	Novel isolated nucleic acid molecule associated with cancerous state of
PT	prostate cells and correlating with presence of prostate cancer, useful
PT	for detecting presence of prostate cancer, stage of prostate cancer.
XX	XX
PS	Claim 1; Page 10673; 11750pp; English.
XX	XX
CC	The invention relates to an isolated nucleic acid molecule (I) comprising
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC	specification or its complement. (I) is useful for: (a) assessing whether
CC	a patient is afflicted with prostate cancer; (b) monitoring the
CC	progression of prostate cancer in a patient; (c) assessing the efficacy
CC	of a test compound to inhibit prostate cancer in a patient; (d) assessing

the efficacy of a therapy for inhibiting prostate cancer in a patient;  
(e) selecting a composition for inhibiting prostate cancer in a patient;  
(f) assessing the prostate cell carcinogenic potential of a compound; (g)  
determining whether prostate cancer has metastasized in a patient; (h)  
assessing the aggressiveness or indolence of prostate cancer in a patient  
; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
Sequence 199 BP; 105 A; 34 C; 37 G; 22 T; 0 U; 1 Other;

CC Oligomers specific to each of the genes are useful for detecting the  
CC methylation state of all CpG dinucleotides within the 350 sequences or  
CC (II) and their complementary sequences, as primer oligonucleotides for  
CC the amplification of the 350 sequences, 'III' and/or their complements and  
CC as oligomer probes for detecting the cytosine methylation state and/or  
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this  
CC patent did not form part of the printed specification but is based on  
CC sequence information supplied to Derwent by the European Patent Office  
XX  
XX Sequence 6036 BP; 1473 A; 98 C; 1516 G; 2949 T; 0 U; 0 Other;

CC the histidine biosynthesis pathway. The sequence can be used to create  
CC transgenic plants which express different amounts of the protein, to  
CC identify loss of function mutants and to produce the protein in a host  
CC cell, for example a bacterium. The protein can be used to identify  
CC inhibitors which may be useful as fungicides and herbicides. The gene was  
CC identified by computer screening for sequences which could encode  
CC histidine biosynthetic enzymes

XX  
SQ Sequence 671 BP; 217 A; 104 C; 144 G; 205 T; 0 U; 1 Other;

Query Match 14.1%; Score 33.6; DB 3; Length 671;  
Best Local Similarity 59.4%; Pred. No. 6.4; Indels 0; Gaps 0;  
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 143 TTGCTCATCAAGTTTACCAAGGACITTTACTATGTAATTAATTTGCAACTAGTAGT 202  
DB 572 TTGTGTAATTTTAAAGGAATATATTTTACCACTGGAGCACTATCTCGACATTTCT 631  
QY 203 CAGATCAATAAATTTAGCTGCGCAAAAAA 238  
DB 632 TATTACAAATTTTCTGTCGCAAAAAA 667

## RESULT 4

ACH29861  
ID ACH29861 standard; cDNA; 404 BP.

XX AC ACH29861;

XX DT 13-OCT-2003 (first entry)

XX XX Human testis cDNA #247.

XX DE Human; ss; sequencing by hybridisation; SH; expressed sequence tag; EST;  
XX KW genome mapping; biodiversity; genetic disorder.

XX OS Homo sapiens.

XX XX US2003073623-A1.

XX PN 17-APR-2003.

XX XX 30-JUL-2001; 2001US-00918995.

XX XX 30-JUL-2001; 2001US-00918995.

XX XX (DRNA/) DRMANAC R T.

XX XX (LABA/) LABAT I.

XX XX (STAC/) STACHE-CRAIN B.

XX XX (DICK/) DICKSON M C.

XX XX (JONE/) JONES L W.

XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX XX WPI; 2003-615964/58.

XX PT New polynucleotide sequences obtained from various cDNA libraries, useful  
XX as hybridization probes, as oligomers for PCR, for chromosome and gene  
XX mapping, in the recombinant production of protein, or in generating  
XX antisense DNA or RNA.

XX FS Claim 1; SEQ ID NO 17073; 44pp; English.

XX CC The invention relates to an isolated polynucleotide comprising any one of  
XX 36243 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
XX determined by the technique of SH (sequencing by hybridisation). Also  
XX included is a purified polypeptide comprising a sequence corresponding to  
XX a reading frame of the novel polynucleotide. The nucleic acid sequences  
XX are useful in diagnostics as expressed sequence tags (EST) for  
XX identifying expressed genes or for physical mapping of the human genome,  
XX in forensics, in assessing biodiversity, or in identifying mutations  
XX responsible for genetic disorders and other traits. The nucleotide  
XX sequences are also useful as hybridisation probes, as oligomers for PCR,

CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX  
SQ Sequence 404 BP; 143 A; 43 C; 77 G; 137 T; 0 U; 4 Other;

Query Match 14.0%; Score 33.4; DB 8; Length 404;  
Best Local Similarity 51.8%; Pred. No. 6.3;  
Matches 100; Conservative 0; Mismatches 91; Indels 2; Gaps 1;

QY 46 TTCTCTCAATGGGCACTGCTTATATCCATTCCTCCCAATCCATTCGAGTTCGAA 105  
DB 161 TTGTGTATATGCACAGAGCTCTTTTATACTAAGCAAAATTAATTTTGTACTAGAA 220  
QY 106 AAGGACTTCCCAAGTTATTCGAGTACTGTGAAGAGTTCTCATGAGTTTACCCAAG 165  
DB 221 AAAAATTGAACTTTTATGTTCTTGGCTTNTAAAAATGTTAATTCAGAA--TTAGTTTAT 278  
QY 166 GACTTTACTATGTGAATTAATTTCTCAACTAGTAGTCAGATCAATAAAATTTACGTGG 225  
DB 279 GCTTTAATTAACCTAATTAATAGCTTTGGACACTTAAAGAGCTCTAAATTTGTTGAA 338  
QY 226 CAAAAA 338  
DB 339 AAAAAA 351

## RESULT 5

AAD-6067/c

ID AAD16067 standard; cDNA; 1208 BP.

XX AC AAD16067;

XX DT 22-NOV-2001 (first entry)

XX DE Soybean zinc transporter (ZnT-2) cDNA from clone se6.pk0012.h2.fis.

XX KW Heavy metal transporter; zinc transporter; transgenic plant; soybean;

XX KW ZnT-2; ss.

XX OS Glycine max.

XX PH Key Location/Qualifiers

XX CDS 1..1050

XX FT /\*tag= a

XX FT /product= "Soybean zinc transporter protein"

XX FT /note= "CDS does not include start codon"

XX FT /partial

XX XX US6278042-B1.

XX XX 21-AUG-2001.

XX XX 14-DEC-1999; 99US-00461474.

XX XX 16-DEC-1998; 98US-0112562P.

XX XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX XX Allen SM, Rafalski JA, Sakai H;

XX XX WPI; 2001-540412/60.

XX XX P-PSDB; AAE09323.

XX PT New arsenic transporter polypeptides and polynucleotides, useful for  
XX producing transgenic plants with altered level of heavy metal tolerance.  
XX PS Example 4; Col 41-42; 31pp; English.

CC The invention relates to heavy metal (e.g. arsenite and zinc) transporter  
CC polypeptides and polynucleotides. The polypeptides of the invention are  
CC useful for producing antibodies that are useful for screening cDNA  
CC expression libraries to isolate full length clones of interest. The  
CC antibodies are also useful for detecting the polypeptides in situ, in  
CC cells or in vitro, in cell extracts. Nucleic acid molecules of the  
CC invention are useful for producing transgenic plants with altered  
CC tolerance to heavy metals. They are also useful as probes for genetically  
CC and physically mapping the genes that are a part of, and as markers for  
CC traits linked to those genes. The present sequence is soybean zinc  
CC transporter (ZnT-2) protein cDNA

XX Sequence 1208 BP; 347 A; 208 C; 281 G; 372 T; 0 U; 0 Other;

Query Match 14.0%; Score 33.4; DB 4; Length 1208;  
Best Local Similarity 54.5%; Pred. No. 8.8;  
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 52 CAATGCGACACCTCTATATCCATTCGCAATCACATTCGGAGTTCGAAAGGAC 111  
DB 864 CAACAGCCCTTCAAGCTTAGTAGCATCTATCTCACGAGGTGTGTCTCCATCAGGAC 805  
QY 112 TCCCAAGTATGGAGTCTGTCGAAGAGTTCGTCATGAGTTTACCCAGGACTTT 171  
DB 804 TCCCAAAATGTTCCGACATGTTGATGTTGTTGTTCCCAACAATCTAGTGAAGATTAG 745  
QY 172 ACT 174  
DB 744 AGT 742

## RESULT 6

ABL33686/c  
ID ABL33686 standard; DNA; 8781 BP.

XX ABL33686;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1659.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antiandemic; cytosine methylation; antiasthmatic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antihemiparetic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
DS.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DB-01032529.

XX 01-SEP-2000; 2000DB-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.

XX Claim 1; SEQ ID NO 1659; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention

XX Sequence 8781 BP; 2407 A; 65 C; 2093 G; 4216 T; 0 U; 0 Other;

Query Match 13.9%; Score 33.2; DB 6; Length 8781;

Best Local Similarity 55.1%; Pred. No. 19;  
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 121 TTATGGAGTACTGGAAGAGTTCGTATCATCAAGTTTACCCAAAGGACTTTACTATGTA 180  
DB 131 TTATAATAAATCTTTATTTACTATATAAATACTTTTATATAATAAATCTTTATTTACTA 132  
QY 181 ATTAATTTCTCAACTAGTCTAGTCAGATCATAAATTTCTAGTGGCAAAAAA 238  
DB 131 ATTATAAACAATATATTAACATATAAACAATAAATACTACCTTCTAAAAAATAA 74

## RESULT 7

ABL16740/c  
ID ABL16740 standard; DNA; 20693 BP.

XX ABL16740;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 1693.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO20011042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

XX Claim 1; SEQ ID NO 1693; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABBS7202). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [fip.wipo.int/pub/published\\_pcr\\_sequences](http://fip.wipo.int/pub/published_pcr_sequences)

XX Sequence 20693 BP; 6280 A; 4195 C; 3675 G; 6543 T; 0 U; 0 Other;



Query Match 13.9%; Score 33; DB 4; Length 20693;  
 Best Local Similarity 49.2%; Pred. No. 28;  
 Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
 QY 54 ATTGGCACTGCTATTATCCATTCGCCATCACTTCGATGCTCTCGAAAGGACTT 113  
 Db 11638 ATTTCCAAATGATTATCTTTGATACGTTACCTTTGGTGTTCCTCATATATAT 11579  
 QY 114 CCCAAGTATTGGAGTACTGTGAAAGAGTTCGTCMGAAGTTTACCCAAAGCACTTTAC 173  
 Db 11578 TCCAAATACATTCCTCAATGTCCTCAAGTGGTTCGGTAATATCCCAATATGATATAC 11519  
 QY 174 TATGCAATTAATGTCGAACTAGTACATCAATCAATCAATCAATCAATCAATCAATCAAT 230  
 Db 11518 CATCCCATTTTAATGAATACTACTTGTCCGGAACACCAACGTTAGATGGTCAA 11462

RESULT 8  
 ABX09781/c  
 ID ABX09781 standard; DNA; 571 BP.  
 AC ABX09781;  
 DT 22-JAN-2003 (first entry)  
 XX M. incognita RNAi molecule #38 useful for nematode control.  
 DE  
 DE RNAi molecule; double-stranded interfering RNA; nematode control;  
 KW RNA mediated interference; mRNA transcript; nematode gene; growth;  
 KW development; parasitism; reproduction; RNAi vector; mRNA translation;  
 KW nematode inhibitor; agricultural industry; anti-nematode; ds.  
 XX  
 OS Meloidogyne incognita.  
 XX  
 XX W0200196584-A2.  
 XX  
 XX 20-DEC-2001.  
 XX  
 XX 12-JUN-2001; 2001WO-US018911.  
 XX  
 XX 12-JUN-2000; 2000US-0210917P.  
 XX  
 XX (AKKA-) AKKADIX CORP.  
 XX  
 XX Mushegiar AR, Taylor CG, Pettelson JS, Eroshkin AM;  
 XX  
 XX WPI; 2002-139714/18.  
 XX  
 XX RNA mediated interference molecule useful for disrupting cellular process  
 in a nematode, for controlling nematodes comprises genetic regulatory  
 sequences.

Claim 1; Page 95; 103pp; English.  
 The present invention relates to RNAi (double-stranded interfering RNA or  
 RNA mediated interference) molecules (nematode genes), and methods of  
 using these sequences in nematode control. RNAi molecules selectively  
 target mRNA transcripts of essential nematode genes. The RNAi molecules  
 of the invention are useful for disrupting cellular processes in a  
 nematode by contacting the nematode with a composition comprising an RNAi  
 molecule. The RNAi molecules are useful for killing nematodes and/or  
 inhibiting their growth, development, parasitism or reproduction and also  
 for the regulation of levels of specific mRNA in nematodes. Genetic  
 regulatory sequences such as promoters, enhancers and terminators can be  
 used in genetic constructs such as RNAi vectors which can be used for  
 nematode control. The RNAi molecules are capable of targeting and  
 reducing (and, in some cases, preventing) the translation of a specific  
 gene product, and can be used to reduce or prevent mRNA translation in  
 any tissue of the nematode because of its ability to cross tissue and  
 cellular boundaries. The RNAi molecule can be contacted with a nematode  
 by soaking, injection, or consumption of a food source containing an RNAi  
 molecule. The RNAi molecules can also be used as an epigenetic factor to  
 prevent the proliferation of subsequent generations of nematodes, to

CC produce nematode inhibitors or RNAi in the plants, and provide new  
 CC biotechnological strategies for managing nematodes under sustainable  
 CC agricultural conditions. ABX09677-ABX09815 represent RNAi molecules  
 CC useful for the control of nematodes  
 XX  
 SQ Sequence 571 BP; 151 A; 142 C; 72 G; 205 T; 0 J; 1 Other;  
 Query Match 13.8%; Score 32.8; DB 6; Length 571;  
 Best Local Similarity 64.5%; Pred. No. 10;  
 Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
 QY 153 AAGGACTTTACTATGTCGAATTAATTCGAACTAGTACATCAATCAATCAATCAATCAAT 222  
 Db 82 AAGGACTTTAAATAAGTGTGTAAATTTTAAATGATTTTCTGATTAATTTATTTGTA 23  
 QY 223 TGGCAAAAAA 238  
 Db 22 CAAAAA 7

RESULT 9  
 ABQ65781/c  
 ID ABQ65781 standard; DNA; 678 BP.  
 XX  
 AC ABQ65781;  
 DT 21-AUG-2002 (first entry)  
 XX Arabidopsis thaliana polynucleotide SEQ ID NO 358.  
 DE  
 DE Arabidopsis thaliana; thale cress; plant; transgenic; GMO; disease;  
 KW stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;  
 KW insecticide; antibiotic; ds.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 XX US2002059663-A1.  
 XX  
 XX 16-MAY-2002.  
 XX  
 XX 26-JAN-2001; 2001US-00770149.  
 XX  
 XX 27-JAN-2000; 2000US-0178506P.  
 XX

(GORL/) GORLACH J.  
 (ANTY/) AN Y.  
 (HAM/) HAMILTON C M.  
 (PRIC/) PRICE J L.  
 (RAIN/) RAINES T M.  
 (YUY/) YU Y.  
 (RAME/) RAMEAKA J G.  
 (PAGE/) PAGE A.  
 (MATH/) MATHW A V.  
 (LEDF/) LEDFORD B L.  
 (WOES/) WOESSNER J P.  
 (HAAS/) HAAS W D.  
 (GARC/) GARCIA C A.  
 (KRICK/) KRICKER M.  
 (SLAT/) SLATER T.  
 (DAVI/) DAVIS K R.  
 (ALLE/) ALLEN K.  
 (HOFF/) HOFFMAN N.  
 (HURB/) HURBAN P.  
 Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,  
 Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;  
 Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;  
 Hurban P;  
 WPI; 2002-479224/51.  
 New nucleic acid that hybridizes to Arabidopsis thaliana sequences,  
 useful e.g. for preparing transgenic plants with increased resistance or

PT altered metabolism.  
XX Claim 1; SEQ ID NO 358; 40pp + Sequence Listing; English.  
PS The invention relates to nucleic acids (I) that hybridize under stringent  
CC conditions to any of 999 sequences (ABQ65424-ABQ66422) or their  
CC fragments. (I) are used to express the corresponding polypeptides (II) or  
CC to produce genetically modified plant cells or transgenic plants, which  
CC may have improved resistance to disease or stress, or altered  
CC metabolic/biosynthetic pathways (for production of commercial,  
CC nutritional or medicinal products), or generally any trait of interest,  
CC or can be used to screen for biologically active agents (e.g. fungicides,  
CC insecticides and antibiotics). Note: The sequence data for this patent  
CC did not form part of the prior specification, but was obtained in  
CC electronic format directly from the USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=999909770143  
XX  
SQ Sequence 678 BP; 161 A; 198 C; 110 G; 209 T; 0 U; 0 Other;  
  
Query Match 13.8%; Score 32.8; DB 6; Length 678;  
Best Local Similarity 64.5%; Pred. No. 11;  
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
  
QY 163 AAGGACTTTACTGTCGATTAATGTCACAACTAGTCAGATCAATCAATTCCTACG 222  
Db |||||  
83 ATGTAATGTTTATGTTGGTTGATGTCATATCTGCTGTTGTCATATAAATTCCTTG 24  
  
QY 223 TGGCAAAAAA 238  
Db |||||  
23 TAGTTAAAAA 8  
  
RESULT 10  
AAK87745/c  
ID AAK87745 standard; CDNA; 745 BP.  
XX  
AC AAK87745;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen coding sequence SEQ ID NO: 61.  
XX  
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
XX digestive system disorder; Meckel's diverticulum; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200155314-A2.  
XX  
ED 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001324.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0195123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 22-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229341P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234232P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246608P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 11-DEC-2000; 2000US-0251990P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-502630/55.  
 DR P-PSDB; AAM91972.  
 XX  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases.  
 XX  
 PS Claim 1; SEQ ID NO 61; 966pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
 CC system antigen of the invention  
 XX  
 SQ Sequence 745 BP; 143 A; 189 C; 167 G; 234 T; 0 U; 12 Other;

Query Match 13.8%; Score 32.8; DB 4; Length 745;  
 Best Local Similarity 60.5%; Pred. No. 11;  
 Matches 52; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 153 AGTTATCCCAAGGAAGCTTACTAATGTAATTAATTTGCAACTAGTAGTCAGATCAATA 212

Db 108 AGATGATTTTAGGACATTTTATTTCATTAATTTTACAATCTAATGTTNANAAAAA 49  
 QY 213 AAATCTACGTGGCAAAAAA 238  
 Db 48 AAAAAA 23  
 RESULT 11  
 AAI57549/c  
 ID AA-57549 standard; cDNA: 745 BP.  
 XX AC AAI57549;  
 XX DT 19-OCT-2001 (first entry)  
 XX DE Human colorectal cancer antigen cDNA SEQ ID NO: 13.  
 XX KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
 XX OS Homo sapiens.  
 XX EN WO200155350-A1.  
 XX PD 02-AUG-2001.  
 XX PF 17-JAN-2001; 2001WO-US001350.  
 XX PR 21-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216800P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218293P.  
 PR 26-JUL-2000; 2000US-0223963P.  
 PR 26-JUL-2000; 2000US-0223964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226686P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 06-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.

```

PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02539678P.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruber SM;
XX
XX WPL; 2001-457727/49.
XX P-PS08; AAM39571.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
XX PT disorders related to the colon and rectum including colorectal cancers
XX PT and also for testing and detection e.g. diagnosis.
XX PT
XX
XX Claim 1; SEQ ID NO 13; 522pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX CC number of colorectal cancer antigens. These are shown in AAI57547-
XX CC AAI57619 and AAM38569-AAM38641. These can be used in the diagnosis,
XX CC prevention and treatment of cancer of the colon and/or rectum. The
XX CC present sequence is a colorectal cancer antigen coding sequence of the
XX CC invention. Note: The sequence data for this patent did not form part of
XX CC the printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 745 BP; 143 A; 189 C; 167 G; 234 T; 0 U; 12 Other;
XX
XX Query Match 13.8%; Score 32.8; DB 4; Length 745;
XX Best Local Similarity 60.5%; Pred.No.11;
XX Matches 52; Conservative 0; Mismatches 34; Indels 0; Gaps 0
XX
XX 153 AGTTTATCCCAAGGACCTTACTATGTAATTAAATTCGCAACTAGTTCAGATCAATA 212
XX 108 AGATGATTTAGGCACATTTTATTTAAATTAATTTACAATCTTAATGTTNANAAAAAAA 49
XX
XX 213 AAATTCACGTGCGCAAAAAAAA 238
XX ||| |
XX 4E AAAAAA
XX
XX ABS99726;
XX
XX 18-DEC-2002 (first entry)
XX
XX cDNA encoding human colorectal cancer related protein #3.
XX
XX Human; colorectal cancer related protein; colon; rectum;
XX KW colorectal cancer metastasis; gastrointestinal disorder; cytostatic;
XX KW gene; ss.
XX
XX Homo sapiens.
XX
XX US2002119919-A1.

```

XX PD 29-AUG-2002.  
 XX PF 17-JAN-2001; 2001US-00764855.  
 XX PR 31-JAN-2000; 2000US-0179065P.  
 XX (ROSE/) ROSEN C A.  
 XX (RUBE/) RUBEN S M.  
 XX (BARA/) BARASH S C.  
 XX Rosen CA, Ruben SM, Barash SC;  
 XX WPI: 2002-731367/79.  
 XX P-PSDB; ABG97623.  
 XX New colorectal cancer polypeptide for diagnosing, prognosing, preventing,  
 PT and treating immune, hyperproliferative, liver, kidney, reproductive  
 PT disorders and for identifying modulators of therapeutic use.  
 XX C:aim 4; SEQ ID NO 13; 183pp; English.  
 XX The present invention relates to the isolation of novel human colorectal  
 CC cancer related proteins, and polynucleotide sequences encoding them. The  
 CC sequences of the invention are useful in the diagnosis, treatment,  
 CC prevention and/or prognosis of the colon and/or rectum, including  
 CC colorectal cancer, colorectal cancer metastases, and gastrointestinal  
 CC disorders such as dysphagia, peptic esophagitis, gastric reflux,  
 CC irritable bowel syndrome, and peritoneal diseases. The invention also  
 CC describes antibodies that bind colorectal cancer related proteins,  
 CC vectors, host cells, and recombinant and synthetic methods for producing  
 CC human colorectal cancer related polynucleotides, polypeptides, and/or  
 CC antibodies. ABG97624-ABG99796 represent cDNA sequences encoding human  
 CC colorectal cancer related proteins. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/psipspIDEntry.html  
 XX Seqdata.uspto.gov/psipspIDEntry.html  
 XX Sequence 745 BP; 143 A; 189 C; 167 G; 234 T; 0 U; 12 Other;  
 Query Watch 13.8%; Score 32.8; DB 6; Length 745;  
 Best Local Similarity 60.5%; Pred. NO. 11;  
 Matches 52; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
 Qy 153 AGTTTACCCAGGACTTTACTATGTCGAAATTAATTCGCAACTAGTACGATCAGATCAATA 212  
 Db 108 AGATGATTTAGGACATTTATTTTAAATTTAAATTTACATCTATGTTGNNAAAAA 49  
 Qy 213 AAATTTCTAGTGGCAAAAAA 238  
 Db 48 AAAAAA 23  
 RESULT 13  
 ADB92806/c  
 ID ADB92806 standard; cDNA; 745 BP.  
 XX AC ADB92806;  
 XX 04-DEC-2003 (first entry)  
 XX Human colorectal cancer related polypeptide cDNA #3.  
 XX ss; gene; human; colorectal cancer; antigen; gene therapy;  
 KW gastrointestinal disorder; inflammatory disease; infection; cancer;  
 KW intestinal neoplasm; small intestine carcinoma; tumour;  
 KW small intestine non-Hodgkin's lymphoma; small bowel lymphoma; ulcer;  
 KW peptic ulcer; Bruton's disease; X linked infantile agammaglobulinaemia;  
 KW severe combined immunodeficiency; DiGeorge anomaly;  
 KW hyperproliferative disorder; acute lymphoblastic leukaemia;  
 KW acute lymphocytic leukaemia; urinary system disorder; cortical necrosis;  
 KW kidney infarction; cardiovascular disorder; carcinoma heart disease;  
 KW arrhythmia; respiratory disorder; non-allergic rhinitis; sinusitis;

KW musculoskeletal system disorder; Albers-Schonberg disease;  
 KW Marfan's syndrome; neurological disease; phenylketonuria;  
 KW Wernicke's encephalopathy; Alzheimer's disease; endocrine disorder;  
 KW Grave's disease; Cushing's syndrome; reproductive system disorder;  
 KW prostatitis; benign prostatic hyperplasia; benign prostatic hyperplasia;  
 KW thrombosis; atherosclerosis; myocardial infarction; ischaemic attack.  
 XX Homo sapiens.  
 OS  
 XX US2003054420-A1.  
 XX 20-MAR-2003.  
 XX 11-FEB-2002; 2002US-00072349.  
 XX 31-JAN-2000; 2000US-0179065P.  
 XX 04-FEB-2000; 2000US-0180628P.  
 XX 24-FEB-2000; 2000US-0184664P.  
 XX 02-MAR-2000; 2000US-0186350P.  
 XX 16-MAR-2000; 2000US-0189874P.  
 XX 17-MAR-2000; 2000US-0190076P.  
 XX 18-APR-2000; 2000US-0198123P.  
 XX 19-MAY-2000; 2000US-020515P.  
 XX 07-JUN-2000; 2000US-0209467P.  
 XX 28-JUN-2000; 2000US-0211486P.  
 XX 30-JUN-2000; 2000US-0215135P.  
 XX 07-JUL-2000; 2000US-0216647P.  
 XX 07-JUL-2000; 2000US-0216880P.  
 XX 11-JUL-2000; 2000US-0217487P.  
 XX 11-JUL-2000; 2000US-0217496P.  
 XX 14-JUL-2000; 2000US-0218290P.  
 XX 26-JUL-2000; 2000US-0220963P.  
 XX 14-AUG-2000; 2000US-0224518P.  
 XX 14-AUG-2000; 2000US-0224519P.  
 XX 14-AUG-2000; 2000US-0225213P.  
 XX 14-AUG-2000; 2000US-0225214P.  
 XX 14-AUG-2000; 2000US-0225266P.  
 XX 14-AUG-2000; 2000US-0225267P.  
 XX 14-AUG-2000; 2000US-0225268P.  
 XX 14-AUG-2000; 2000US-0225447P.  
 XX 14-AUG-2000; 2000US-0225447P.  
 XX 14-AUG-2000; 2000US-0225757P.  
 XX 14-AUG-2000; 2000US-0225758P.  
 XX 18-AUG-2000; 2000US-0225759P.  
 XX 22-AUG-2000; 2000US-0226681P.  
 XX 22-AUG-2000; 2000US-0226688P.  
 XX 22-AUG-2000; 2000US-0227182P.  
 XX 23-AUG-2000; 2000US-0227009P.  
 XX 30-AUG-2000; 2000US-0228924P.  
 XX 01-SEP-2000; 2000US-0229287P.  
 XX 01-SEP-2000; 2000US-0229343P.  
 XX 01-SEP-2000; 2000US-0229344P.  
 XX 01-SEP-2000; 2000US-0229345P.  
 XX 05-SEP-2000; 2000US-0229509P.  
 XX 06-SEP-2000; 2000US-0229513P.  
 XX 06-SEP-2000; 2000US-0230437P.  
 XX 08-SEP-2000; 2000US-0230438P.  
 XX 08-SEP-2000; 2000US-0231242P.  
 XX 08-SEP-2000; 2000US-0231243P.  
 XX 08-SEP-2000; 2000US-0231244P.  
 XX 08-SEP-2000; 2000US-0231413P.  
 XX 08-SEP-2000; 2000US-0231414P.  
 XX 08-SEP-2000; 2000US-0232080P.  
 XX 12-SEP-2000; 2000US-0231968P.  
 XX 14-SEP-2000; 2000US-0232397P.  
 XX 14-SEP-2000; 2000US-0232398P.  
 XX 14-SEP-2000; 2000US-0232399P.  
 XX 14-SEP-2000; 2000US-0232400P.  
 XX 14-SEP-2000; 2000US-0232401P.  
 XX 14-SEP-2000; 2000US-0233063P.

[illegible]

PR	38-DEC-2000; 2000US-0251989P.
PR	08-DEC-2000; 2000US-0251990P.
PR	11-DEC-2000; 2000US-0254097P.
PR	05-JAN-2001; 2001US-0259678P.
PR	17-JAN-2001; 2001US-0075485P.
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
XX	Rosen CA, Ruben SM, Barash SC;
XX	WPI; 2003-708345/67.
DR	P-PSDB; ADB2879.
DR	
PT	Novel colorectal cancer antigen useful for treating, preventing,
PT	diagnosing and/or proposing gastrointestinal disorders, infections,
PT	cancers such as intestinal neoplasms, ulcers.
XX	
XX	Claim 3; SEQ ID NO 13; 179pp; English.
XX	
CC	The invention relates to a colorectal cancer antigen. The antigen is
CC	useful for chromosome identification, chromosome mapping, radiation
CC	hybrid mapping or gene therapy, or as hybridisation probes for
CC	differential identification of the tissues or cell types present in a
CC	biological sample. The antigen is useful for treating, preventing,
CC	diagnosing and/or proposing gastrointestinal disorders, including
CC	inflammatory diseases and/or conditions, infections, cancers (e.g.
CC	intestinal neoplasms (carcinoid tumour of the small intestine, not-
CC	Hodgkin's lymphoma of the small intestine, small bowel lymphoma)) and
CC	ulcers (e.g. peptic ulcers). The antigen and its nucleic acid are useful
CC	to provide immunological probes for differential identification of the
CC	tissue. The antigen and its nucleic acid are useful for treating,
CC	preventing, diagnosing and/or prognosing diseases, disorders and/or
CC	conditions of the immune system e.g. Bruton's disease, X linked infantile
CC	agammaglobulinaemia, severe combined immunodeficiencies, DiGeorge
CC	anomaly, etc. The antigen and its nucleic acid is useful for treating,
CC	preventing and/or diagnosing hyperproliferative disorders (e.g. acute
CC	
Query Match	13.8%; Score 32.8; DB 9; Length 745;
Best Local Similarity	60.5%; Pred. No. 11;
Matches 52; Conservative 0; Mismatches 34; Indels 0; Gaps 0;	
Qy	153 AGTTTACCAGGAGCTTTACTATGTCGATTAAATGTCAACTAGTTCAGATCAATA 212
Db	108 AGATGATTTAGGACATTTATTATTAAATTACATCTAATGTTNANAAAAA 49
Qy	213 AAATTCCTACGTGGCAAAAAA 238
Db	48 AAAAAA 23
RESULT 14	
ABL29872	
ID	ABL29872 standard; DNA; 2630 BP.
XX	
AC	ABL29872;
XX	
DC	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 41089.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ds.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US009231.
XX	
PR	23-MAR-2000; 2000US-C191637P.
PR	11-JUL-2000; 2000US-C0614150.

```
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers BW;
XX DR WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX
XX Claim 1; SEQ ID NO 41089; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2630 BP; 725 A; 579 C; 613 G; 713 T; 0 U; 0 Other;
XX
Query Match 13.8%; Score 32.8; DB 4; Length 2630;
Best Local Similarity 49.4%; Pred. No. 17;
Matches 55; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 23 TCACGGGGGGGATGCAACAGCTCTCTCAATTGGCAACTGCTTATTCATTCCCGCA 82
Db 505 TAATGCCAGTTAAAGCTTAAGATGTTTGTCAAGTGGTGTCATAAATTATATGTTATCGT 564
QY 83 ATCACATTTCGGATGTTCTCGAAAGAGCTTCCCAAGTATTGGAGTACTGTGAAAGNG 142
Db 565 CTATCAGTATGATGATGCTTAAAGATGTTGTTGTCATTTCATTGAGTTAAAGTTAAAAA 624
QY 143 TTCGTCATGGAAGTTTACCCAAAGGACTTCTATGTAATTAATTTGCAAA 194
Db 625 TTGTAGCTTTTGGCCCTCTGCAATGTAAGATATATTCACGGGCAAA 676
RESULT 15
ABQ708-9
ID ABQ70819 standard; DNA; 3423 BP.
XX AC ABQ70819;
XX
XX 29-AUG-2003 (revised)
XX DT 29-AUG-2002 (first entry)
XX
XX Listeria monocytogenes 4b contig DNA sequence #761.
XX
XX Antibacterial; Listeria; food contamination; mutational analysis;
XX XW infection; ds.
XX
XX Listeria monocytogenes ATCC 19115.
XX
XX WO200228891-A2.
XX
XX 11-APR-2002.
XX
XX 04-OCT-2001; 2001WO-FR003061.
XX
XX 04-OCT-2000; 2000FR-0002697.
XX
XX (INSP ) INST PASTEUR.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Kunst F, Glaser P;
XX
XX WPI; 2002-332479/37.
XX
XX New genomic sequences from Listeria species, useful for detection,
XX PT treatment and prevention of infection, also related polypeptides,
XX PT antibodies and modulators.
XX
XX Claim 14; SEQ ID NO 3632; 180pp; French.
XX
XX The present invention relates to nucleic acid sequences (ABQ67188-
XX CC ABQ71212) from Listeria sp. The sequences are useful as probes and
XX CC primers for identification and/or detection of Listeria (e.g. as
XX CC contaminants in foods, or mutational analysis) and for analysis of gene
XX CC expression. Proteins encoded by the nucleic acid sequences can be used to
XX CC screen for compounds that modulate gene expression, replication and
XX CC pathogenicity of Listeria (potential therapeutic agents), also for
XX CC treating infections by Listeria, and are useful as immunogens in anti-
XX CC Listeria vaccines. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
XX CC on 29-AUG-2003 to standardise OS field)
XX
XX Sequence 3423 BP; 1040 A; 693 C; 620 G; 1070 T; 0 U; 0 Other;
XX
Query Match 13.8%; Score 32.8; DB 6; Length 3423;
Best Local Similarity 59.8%; Pred. No. 19;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 58 GCAACTGTCTATTATCCATTCGCAATTCACATTTGGGATGTTCTCGAAGAGGACTCCCA 117
Db 2377 GAATTGTTTATCTTTTCCCAAGCAAAAGCAATGATGATGATGATGATGATGATGATGAT 2436
QY 118 AAGTTATGAGTACTGTGAAAGAGTTCGTCA 149
Db 2437 CAACATCTGGCATAGTACGAATGTTTCGCA 2468
Search completed: July 7, 2004, 15:07:29
Job time : 353 secs
```







```
Db 3900 TATTATGCGCAGATGGAGCTTGTGAAGACACTTCAAGTATTCCTTGT 3953
RESULT 5
US-08-213-419B-3/c
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: J11-002CNC
; CURRENT APPLICATION NUMBER: US/08/213,419B
; PRIOR FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6:24
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3
Query Match 13.2%; Score 31.4; DB 4; Length 6124;
Best Local Similarity 47.3%; Pred. No. 4.3; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

QY 38 CAACAGCTCTCTCAATGGCAACTGTCTATTATCCATTCCGCAATCACAATTCGGAG 97
DB 865 CATGAATACCTTCCATTCTGTAATGGAATTTATTTTCCATAATGATATATAATA 806
QY 98 TTCTCGAAGAGACTCCCAAGTTATTTGGAGTACTGTGAAGAGTTCGTCAAGATTT 157
DB 805 TATATATTAAAGCAAAATAGATATTAAATAAAGAAATACATAATAGTTTAA 746
QY 158 ACCCAAGAGACTTTACTATGTAATTAATTCCTCAACTAGTAGTCAGATCAATAAAT 217
DB 745 CTAAAGACTAGTTTATTTCTATTTTAAATTTTAAATGTAATGTAATGTAATGTAAT 686
QY 218 CTACGTGGCAAAAAA 238
DB 685 TAAATAAATAAATAAATAA 665
RESULT 6
US-09-302-812-38/c
; Sequence 38, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; PRIOR FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
```

```
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-302-812-38
Query Match 13.2%; Score 31.4; DB 4; Length 29793;
Best Local Similarity 51.0%; Pred. No. 7.8; Indels 0; Gaps 0;
Matches 74; Conservative 0; Mismatches 71;

QY 94 GATGTTCTCGAAAAGGACTTCCCAAAGTTATTGGAGTACTGTGAAGAGTTTCGTCATGAA 153
DB 21188 GATTTCTACATTCCTTATTCGCAACGATGTTTCAATATGTTATGAAATTTCTGTT 21129
QY 154 GTTTACCCCAAGGACITTTACTATGTGAATTAATGTCAAATCTAGTCTAGATCAATAA 213
DB 21128 TTTTCCCATTAAGTGAATTCAAAATATTTATTGTCATATGCTTGTTCTAGTAAT 21069
QY 214 AATCTAGTCGCAAAAAA 238
DB 21068 TTTAATTTTGAATAAAAAA 21044
RESULT 7
US-09-511-477-38/c
; Sequence 38, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG)
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACT
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-511-477-38
Query Match 13.2%; Score 31.4; DB 4; Length 29793;
Best Local Similarity 51.0%; Pred. No. 7.8; Indels 0; Gaps 0;
Matches 74; Conservative 0; Mismatches 71;

QY 94 GATGTTCTCGAAAAGGACTTCCCAAAGTTATTGGAGTACTGTGAAGAGTTTCGTCATGAA 153
DB 21188 GATTTCTACATTCCTTATTCGCAACGATGTTTCAATATGTTATGAAATTTCTGTT 21129
QY 154 GTTTACCCCAAGGACITTTACTATGTGAATTAATGTCAAATCTAGTCTAGATCAATAA 213
DB 21128 TTTTCCCATTAAGTGAATTCAAAATATTTATTGTCATATGCTTGTTCTAGTAAT 21069
QY 214 AATCTAGTCGCAAAAAA 238
DB 21068 TTTAATTTTGAATAAAAAA 21044
RESULT 8
US-09-511-507-38/c
; Sequence 38, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
```

```

1 EARLIER FILING DATE: 1397-07-08
2 EARLIER APPLICATION NUMBER: 60/051,919
3 EARLIER FILING DATE: 1397-07-08
4 EARLIER APPLICATION NUMBER: 60/051,928
5 EARLIER FILING DATE: 1397-07-08
6 EARLIER APPLICATION NUMBER: 60/055,722
7 EARLIER FILING DATE: 1397-08-18
8 EARLIER APPLICATION NUMBER: 60/055,723
9 EARLIER FILING DATE: 1397-08-18
10 EARLIER APPLICATION NUMBER: 60/055,948
11 EARLIER FILING DATE: 1397-08-18
12 EARLIER APPLICATION NUMBER: 60/055,949
13 EARLIER FILING DATE: 1397-08-18
14 EARLIER APPLICATION NUMBER: 60/055,953
15 EARLIER FILING DATE: 1397-08-18
16 EARLIER APPLICATION NUMBER: 60/055,950
17 EARLIER FILING DATE: 1397-08-18
18 EARLIER APPLICATION NUMBER: 60/055,947
19 EARLIER FILING DATE: 1397-08-18
20 EARLIER APPLICATION NUMBER: 60/055,964
21 EARLIER FILING DATE: 1397-08-18
22 EARLIER APPLICATION NUMBER: 60/056,360
23 EARLIER FILING DATE: 1397-08-18
24 EARLIER APPLICATION NUMBER: 60/055,684
25 EARLIER FILING DATE: 1397-08-18
26 EARLIER APPLICATION NUMBER: 60/055,984
27 EARLIER FILING DATE: 1397-08-18
28 EARLIER APPLICATION NUMBER: 60/055,954
29 EARLIER FILING DATE: 1397-08-18
30 EARLIER APPLICATION NUMBER: 60/058,785
31 EARLIER FILING DATE: 1397-09-12
32 EARLIER APPLICATION NUMBER: 60/058,664
33 EARLIER FILING DATE: 1397-09-12
34 EARLIER APPLICATION NUMBER: 60/058,660
35 EARLIER FILING DATE: 1397-09-12
36 EARLIER APPLICATION NUMBER: 60/058,661
37 EARLIER FILING DATE: 1397-09-12
38 NUMBER OF SEQ ID NOS: 672
39 SOFTWARE: PatentIn Ver. 2.0
40 SEQ ID NO 69
41 LENGTH: 696
42 TYPE: DNA
43 ORGANISM: Homo sapiens
44 FEATURE:
45 NAME/KEY: SITE
46 LOCATION: [605]
47 OTHER INFORMATION: n equals a,t,g, or c
48 FEATURE:
49 NAME/KEY: SITE
50 LOCATION: [648]
51 OTHER INFORMATION: n equals a,t,g, or c
52 FEATURE:
53 NAME/KEY: SITE
54 LOCATION: [655]
55 OTHER INFORMATION: n equals a,t,g, or c
56 US-09-227-357-69
57
58 Query Match 13.1%; Score 31.2; DB 4; Length 696;
59 Best Local Similarity 50.7%; Pred.No.2.2;
60 Matches 72; Conservative 1; Mismatches 69; Indels 0; Gaps 0;
61
62 QY 95 ATGTTCTCGAAAGGACTTCCTCCAAAGTATTGAGTACTCTGAAAGAGTTCGTTCATGAAG 154
63 DB 278 ATGCTTTTAAAATCCWAACACTACTATATCAGTACCATGAGTCAICGCGTTGAGG 219
64
65 QY 155 TTTACCCAAAGGACTTTACTATGFGAATTAATTTGTCAAACTAGTAGTCAGATCAATAAA 214
66 DB 218 GTAACTAATCCCTTATACCACTGATTTATTTTATTATTAGTTAATTAATGACGATTAT 159
67
68 QY 215 ATTCTACGTGGCAAAAAAAA 236
69 DB 158 TTTCACCTGCAAAAAGAAAAA 137

```

```
RESULT 10
US-09-601-198-149
Sequence 149, Application US/09601198
Patent No. 6531583
GENERAL INFORMATION:
APPLICANT: Cassell, Gail H.
APPLICANT: Chen, Eilsoo Y.
APPLICANT: Glass, Jennifer S.
APPLICANT: Heiner, John I.
APPLICANT: Heiner, Cheryl R.
APPLICANT: Lefkowitz, Elliot
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA
FILE REFERENCE: UAB-13452/22
CURRENT APPLICATION NUMBER: US/09/601,198
CURRENT FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/073,189
PRIOR FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 149
LENGTH: 2244
TYPE: DNA
ORGANISM: Ureaplasma urealyticum
US-09-601-198-149

Query Match      12.9%; Score 30.8; DB 4; Length 2244;
Best Local Similarity 55.7%; Pred. No. 4.5;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 90 TTCGATGTTTCGMAAGGACTCCAAAGTATTGGAGTACTGTGAAGAGTTCGTCA 149
Db 338 TTACATTTACATTAACACCACTAAACAAAGTTATTAAATTAATACAGCTGTTTT 397

Qy 150 TGAAGTTTACCAAGAGCTTTACTATGTAATTAATTTGCAAC 195
Db 398 TAACAGATAACAAACAATATCTACAGAAATTAATATAAATATAAAC 443

RESULT 11
US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729
FILE REFERENCE: jannaschii
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148348)..(148348)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231380)..(231380)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
```

/ LOCATION: (657081)..(657081)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (657203)..(657203)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (674335)..(674335)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (682442)..(682442)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (713652)..(713652)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (741684)..(741684)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (779455)..(779455)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (779676)..(779676)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (855539)..(855539)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (871619)..(871619)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1084830)..(1084830)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1096846)..(1096846)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1119881)..(1119881)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1130881)..(1130881)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1310988)..(1310988)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1313224)..(1313224)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1349473)..(1349473)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1349491)..(1349491)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1470091)..(1470091)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1569020)..(1569020)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1602912)..(1602912)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1603734)..(1603734)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1637998)..(1637998)  
 / OTHER INFORMATION: n equals a, t, c, or g  
 / NAME/KEY: misc feature  
 / LOCATION: (1664854)..(1664855)  
 / OTHER INFORMATION: n equals a, t, c, or g

US-09-437-450a-40.rni

Query Match 12.9%; Score 30.6; DB 4; Length 1664976;  
 Best Local Similarity 45.6%; Pred. No. 56; Mismatches 129; Indels 0; Gaps 0;  
 Matches 108; Conservative 0;  
 QY 1 TTTTCTTTTCTTTTGGGAGGAAATCAGCGGCGGGATGCAACAGTCTCTCTCAATGGCA 60  
 DB 619908 TTATAATTACTGACATAGATACCTGTGATATTTAAATAGGACTTTCCGAGAGTAA 619849  
 QY 61 ACTGTCTATTATCCATTCGGCAATCACHTTGGAGTTCTCGAAGAGACTTCCCAAG 120  
 DB 619848 ATTTTCTATTCTTTTATTATCATTTATGATAGGACGTAATTTGATATGATGATCAAAAG 619789  
 QY 121 TTATTCGAGTACTGCGAAGAGTTGTCATGAGTTTACCCAAAGGACTTTTACTATGTA 180  
 DB 619788 GTGTTAAAGTCCAAATAGACCTTTTGTAGTTTAAAGCATTTAGTTCACCTTTTGTGC 619729  
 QY 181 ATTAAATTTGCAAACTAGTACTGAGATCAATATAAATTTCTAGTGGCAAAAAA 237  
 DB 619728 AATGATAAATTTTGTCTACGAACTTCTATCAGATGATATTTATAAAACTACA 619672  
 RESULT 12  
 US-09-437-597-266/c  
 / Sequence 266, Application US/09643597  
 / Patent No. 6426072  
 / GENERAL INFORMATION:  
 / APPLICANT: Wang, Tongtong  
 / APPLICANT: Fan, Liqun  
 / APPLICANT: Kalos, Michael D.  
 / APPLICANT: Bangur, Chaitanya S.  
 / APPLICANT: Hosken, Nancy  
 / APPLICANT: Fanger, Gary R.  
 / APPLICANT: Li, Samuel X.  
 / APPLICANT: Wang, Aijun  
 / APPLICANT: Skeiky, Yasir A.W.  
 / APPLICANT: Henderson, Robert A.  
 / APPLICANT: McNeill, Patricia D.  
 / TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 / FILE REFERENCE: 210121.455C11  
 / CURRENT APPLICATION NUMBER: US/09/643,597  
 / CURRENT FILING DATE: 2000-08-21  
 / NUMBER OF SEQ ID NOS: 369  
 / SOFTWARE: FastSeq for Windows Version 3.0  
 / SEQ ID NO 266  
 / LENGTH: 401  
 / TYPE: DNA  
 / ORGANISM: Homo sapien  
 / FEATURE:  
 / NAME/KEY: misc feature  
 / LOCATION: (1)-(401)  
 / OTHER INFORMATION: n = A,T,C or G  
 US-09-437-597-266

Query Match 12.8%; Score 30.4; DB 4; Length 401;  
 Best Local Similarity 49.7%; Pred. No. 3.1;  
 Matches 76; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
 QY 83 ATCACTTCGGATGTTCTCGAAGAGACTTCCCAAGTATTGAGTACTGTGAAGAG 142  
 DB 181 ATTACAAAATGATTGTTATTTAAAGAAATTTTAAATCCGAAAGTCACTTTAAATAG 122  
 QY 143 TTCGTCATGAGTTTACCAAGGACTTTACTATGTAATTAATTTCTCAAACTAGTAGT 202  
 DB 121 AACCTCATATAGTATGACAACTATAAAATTTTACATTCNUGAANAATTTTCTGT 62  
 QY 203 CAGATCAATAAAATTTCTAGTGGCAAAAAA 235  
 DB 61 CGTCTCATTTATTCNCTGATACAAATGAA 29

US-09-437-597-266/c

; Sequence 266, Application US/0948084A

; Patent No. 6482597

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Fan, Liqun

; APPLICANT: Hosken, Nancy A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.455C6

; CURRENT APPLICATION NUMBER: US/09/480,884A

; CURRENT FILING DATE: 2001-08-27

; NUMBER OF SEQ ID NOS: 330

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 266

; LENGTH: 401

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(401)

; OTHER INFORMATION: n = A,T,C or G

US-09-480-884A-266

Query Match 12.8%; Score 30.4; DB 4; Length 401;

Best Local Similarity 49.7%; Pred. No. 3.1;

Matches 76; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 83 ATCAGATTCGGATGTTCTCGAAAAGGACTTCCCAAGTTATTCGAGTACTGTGAAAGAG 142

DB 181 ATTACAAAATGATTGTTATTTAAAGAAATTTTAAATCCAGAAAGTCATTTAAATAG 122

QY 143 TTGTCATGAGTTTACCCAAAGGACTTTACTATGTAATTAATTTGCAAACTAGTAGT 202

DB 121 AACCTCATATAGTAGAACAACTATAAAATATTTTCAATTCATGAATAATTTTGTGT 62

QY 203 CAGATCAATAAAATTTCTACGTGCGCAAAAAAAA 235

DB 61 CGTCTCATTTATATTCNCTGTATACAAATTGAA 29

RESULT 14

US-09-542-615A-266/c

; Sequence 266, Application US/095426-5A

; Patent No. 6518256

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Fan, Liqun

; APPLICANT: Kalos, Michael D.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Hosker, Nancy A.

; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.455C8

; CURRENT APPLICATION NUMBER: US/09/542,615A

; CURRENT FILING DATE: 2000-04-14

; NUMBER OF SEQ ID NOS: 350

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 266

; LENGTH: 401

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(401)

; OTHER INFORMATION: n = A,T,C or G

US-09-542-615A-266

Query Match

12.8%; Score 30.4; DB 4; Length 401;

Best Local Similarity 49.7%; Pred. No. 3.1;

Matches 76; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 83 ATCAGATTCGGATGTTCTCGAAAAGGACTTCCCAAGTTATTCGAGTACTGTGAAAGAG 142

DB 181 ATTACAAAATGATTGTTATTTAAAGAAATTTTAAATCCAGAAAGTCATTTAAATAG 122

QY 143 TTGTCATGAGTTTACCCAAAGGACTTTACTATGTAATTAATTTGCAAACTAGTAGT 202

DB 121 AACCTCATATAGTAGAACAACTATAAAATATTTTCAATTCATGAATAATTTTGTGT 62

QY 203 CAGATCAATAAAATTTCTACGTGCGCAAAAAAAA 235

DB 61 CGTCTCATTTATATTCNCTGTATACAAATTGAA 29

RESULT 15

US-09-606-421B-266/c

; Sequence 266, Application US/09606421B

; Patent No. 6531315

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Fan, Liqun

; APPLICANT: Kalos, Michael D.

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Hosken, Nancy

; APPLICANT: Fanger, Gary R.

; APPLICANT: Li, Samuel X.

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.455C9

; CURRENT APPLICATION NUMBER: US/09/606,421B

; CURRENT FILING DATE: 2000-06-28

; NUMBER OF SEQ ID NOS: 358

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 266

; LENGTH: 401

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(401)

; OTHER INFORMATION: n = A,T,C or G

US-09-606-421B-266

Query Match 12.8%; Score 30.4; DB 4; Length 401;

Best Local Similarity 49.7%; Pred. No. 3.1;

Matches 76; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 83 ATCAGATTCGGATGTTCTCGAAAAGGACTTCCCAAGTTATTCGAGTACTGTGAAAGAG 142

DB 181 ATTACAAAATGATTGTTATTTAAAGAAATTTTAAATCCAGAAAGTCATTTAAATAG 122

QY 143 TTGTCATGAGTTTACCCAAAGGACTTTACTATGTAATTAATTTGCAAACTAGTAGT 202

DB 121 AACCTCATATAGTAGAACAACTATAAAATATTTTCAATTCATGAATAATTTTGTGT 62

QY 203 CAGATCAATAAAATTTCTACGTGCGCAAAAAAAA 235

DB 61 CGTCTCATTTATATTCNCTGTATACAAATTGAA 29

Search completed: July 7, 2004, 16:22:46

Job time : 87 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 7, 2004, 15:01:33 ; Search time 344 Seconds  
(without alignments)  
3337.679 Million cell updates/sec

Title: US-09-437-450a-40  
Perfect score: 238  
Sequence: 1 tttttttttttgggggga.....taoatggcaaaaaa 238

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3163042 seqs, 241210380 residues

Total number of hits satisfying chosen parameters: 5326084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubna/US09\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubna/US09\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubna/US10\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubna/US10\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubna/US10\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	37	15.5	263	13	US-10-085-783A-49714
2	37	15.5	263	16	US-10-085-783A-49714
3	36.2	15.2	4233	12	US-10-041-018-135
4	34.2	14.4	308	13	US-10-424-599-87084
5	34.2	14.4	614	13	US-10-027-632-94282
6	34.2	14.4	614	13	US-10-027-632-94282
7	34.2	14.4	614	16	US-10-027-632-94282
8	34.2	14.4	614	16	US-10-027-632-94282
9	33.4	14.0	404	10	US-09-918-995-17073
10	33.4	14.0	499	9	US-09-783-590-5790
11	33.2	13.9	8781	15	US-10-311-455-1659
12	32.8	13.8	678	9	US-09-770-149-358
13	32.8	13.8	745	9	US-09-764-855-13
14	32.8	13.8	745	15	US-10-072-349-13

Sequence 3632, Ap  
Sequence 43, Appl  
Sequence 765, App  
Sequence 765, App  
Sequence 440, App  
Sequence 58, Appl  
Sequence 61, Appl  
Sequence 79, Appl  
Sequence 96, Appl  
Sequence 224261,  
Sequence 224261,  
Sequence 4562, Ap  
Sequence 4562, Ap  
Sequence 187, App  
Sequence 289, App  
Sequence 5, Appl  
Sequence 5635, Ap  
Sequence 18833, A  
Sequence 589, App  
Sequence 37, Appl  
Sequence 45512, A  
Sequence 873, App  
Sequence 873, App  
Sequence 30, Appl  
Sequence 1294, Ap  
Sequence 2147, Ap  
Sequence 20058, A  
Sequence 42107, A  
Sequence 44178, A  
Sequence 121, App  
Sequence 38, Appl

## ALIGNMENTS

### RESULT 1

US-10-085-783A-49714  
; Sequence 49714, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ IDS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 49714  
; LENGTH: 263  
; TYPE: DNA  
; ORGANISM: Human  
US-10-085-783A-49714

Query Match 15.5%; Score 37; DB 13; Length 263;  
Best Local Similarity 52.8%; Pred. No. 0.77;  
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
QY 56 TGGCACTGCTATTTCATTCATTCGCAATCACATTCGGATTCCTCGAAAGACTCC 115  
DB 49 TGGTAAAGATTGATTAGTTACCTCATTTTCTTAATGAGTTTAAAGTTTG 108  
QY 116 CAAAGTATTGAGTACTGTGAAGAGTTGCTCATGAGTTTACCCAAAGACTTTACTA 175  
DB 109 CTTTCCCTTGATAGATTGTACATCATGCTCCAGAAATATATTCATTAATTACTA 168

QY 176 TGTGAATTAATTTGTCACAACTAGTAGTCAGATCAATA 212  
DB 169 TAAAGTAATTTTAAACACTTATGTTAAATTACA 205

## RESULT 2

US-10-042-535A-49714  
; Sequence 49714, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liaw, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 49714  
; LENGTH: 263  
; TYPE: DNA  
; ORGANISM: Human  
US-10-042-535A-49714

Query Match 15.5%; Score 37; DB 16; Length 263;  
Best Local Similarity 52.2%; Pred. No. 0.77;  
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 56 TGGCACTGCTTAATTCACATTCGGATTCGATTCGAAAGGACTTCC 115  
DB 49 TGGTAAGATTTGATTTAGTTACTTCATTCTTAATGGAGTTTAAGTAAAGTTG 108  
QY 116 CAACTATTGCGAGTACTGGAAGAGTGTCTGCTCACTTTACCCAAAGGACTTACTA 175  
DB 109 CTTTGCCTTGATGAATTTGTAACATCACTGTCGAAAGATTAATTCATTAATTACTA 169  
QY 176 TGTGAATTAATTTGTCACAACTAGTAGTCAGATCAATA 212  
DB 169 TAAAGTAATTTTAAACACTTATGTTAAATTACA 205

## RESULT 3

US-10-041-018-135  
; Sequence 135, Application US/10041018  
; Publication No. US20040072323A1  
; GENERAL INFORMATION:  
; APPLICANT: Matsuda, Seiichi P.T.  
; APPLICANT: Hart, Elizabeth A.  
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism  
; FILE REFERENCE: P02080U1/10025547  
; CURRENT APPLICATION NUMBER: US/10/041,018  
; CURRENT FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: US 60/259880  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 413  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 135  
; LENGTH: 4233  
; TYPE: DNA  
; ORGANISM: *Gossypium hirsutum*  
US-10-041-018-135

Query Match 15.2%; Score 36.2; DB 12; Length 4233;  
Best Local Similarity 55.0%; Pred. No. 4.8;  
Matches 71; Conservative 0; Mismatches 58; Indels 0;

QY 110 ACTTCCCAAGTTATTGGAGTACTGTGAAGAGTTTCGTATGAAGTTTACCCAAAGGACT 169  
DB 236 ACTTCCAGTTTTCATAAAGATAGGATGGAATTTCTAAATTAACCTTTTAACTCATA 295  
QY 173 TTACTATGTAATTAATTTGTCACAACTAGTAGTCAGATCAATAAATTTCTAGTGGCAA 229  
DB 296 TAAGTAATGGATTAAATTTCCGAAATTAATAAACAGAAAAATTAATTTCAAAAGTATGA 355  
QY 230 AAAAAAAA 238  
DB 356 AATAAAAA 364

## RESULT 4

US-10-424-599-87084  
; Sequence 87084, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yonowei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21/532231B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 87084  
; LENGTH: 308  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_49647C.1  
US-10-424-599-87084

Query Match 14.4%; Score 34.2; DB 13; Length 308;  
Best Local Similarity 60.0%; Pred. No. 5.6;  
Matches 57; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 144 TCGTCATGAGTTTACCCAAAGGACTTTACTATGTGAATTAATTTCTCAAACTAGTAGTC 203  
DB 57 TTGTGATGTTGTTAAGAAATATATTTTACCATGTGCGAGCACTATCTCTGACATTGTT 116  
QY 204 AGATCAATAAATTTCTACGTGGCAAAAAA 238  
DB 117 ATTACAAATTTATTCATGTGTCAAAAAA 151

## RESULT 5

US-10-027-632-94282/c  
; Sequence 94282, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28



```
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94282
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-94282
```

```
Query Match 14.4%; Score 34.2; DB 13; Length 614;
Best Local Similarity 64.6%; Pred. No. 7.7; Mismatches 0; Gaps 0;
Matches 51; Conservative 0; Indels 28; Indels 0; Gaps 0;

QY 160 CCAAGGACTTTACTATGTGAATTAATTTGCAAACTAGTAGTCAGATCAATAAATCT 219
Db 180 CCAATGTAGTCTTAATGTAAATTAATAAATAAATAAATAAATAAATAAATAAAT 121

QY 220 ACCTGGCAAAAAA 238
Db 120 ATGTGACACAAAACAGA 102
```

## RESULT 6

```
US-10-027-632-305384/c
; Sequence 305384, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 305384
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-305384

```
Query Match 14.4%; Score 34.2; DB 13; Length 614;
Best Local Similarity 64.6%; Pred. No. 7.7;
Matches 51; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 160 CCAAGGACTTTACTATGTGAATTAATTTGCAAACTAGTAGTCAGATCAATAAATCT 219
Db 180 CCAATGTAGTCTTAATGTAAATTAATAAATAAATAAATAAATAAATAAATAAAT 121

QY 220 ACCTGGCAAAAAA 238
Db 120 ATGTGACACAAAACAGA 102
```

## RESULT 7

```
US-10-027-632-94282/c
; Sequence 94282, Application US/10027632
; Publication No. US20030204075A9
```

## GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94282
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-94282
```

## Query Match

```
14.4%; Score 34.2; DB 16; Length 614;
Best Local Similarity 64.6%; Pred. No. 7.7;
Matches 51; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
```

```
QY 160 CCAAGGACTTTACTATGTGAATTAATTTGCAAACTAGTAGTCAGATCAATAAATCT 219
Db 180 CCAATGTAGTCTTAATGTAAATTAATAAATAAATAAATAAATAAATAAAT 121

QY 220 ACCTGGCAAAAAA 238
Db 120 ATGTGACACAAAACAGA 102
```

## RESULT 8

```
US-10-027-632-305384/c
; Sequence 305384, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 305384
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-305384

Query Vatch 14.4%; Score 34.2; DB 16; Length 614;  
Best Local Similarity 64.6%; Pred. No. 7.7; Mismatches 0; Gaps 0;  
Matches 51; Conservative 0; Indels 28; Gaps 0;  
QY 160 CCAAGGACTTTACTATGTAATTAANTGTCAAACCTAGTACATCAATAAATTCCT 219  
DB 180 CCCAATGTCCTTAATGTTAATTAATAATATAAGAGGAGGCGAGAGTAATAATAT 121  
QY 220 ACCTGCAAAAAA 238  
DB 120 ACCTGCAAAAAA 102

RESULT 9

US-09-918-995-17073  
; Sequence 17073, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCES: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIORITY FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIORITY FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17073  
; LENGTH: 404  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(404)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-17073

Query Match 14.0%; Score 33.4; DB 10; Length 404;  
Best Local Similarity 51.8%; Pred. No. 11; Mismatches 91; Indels 2; Gaps 1;  
Matches 100; Conservative 0; Indels 2; Gaps 1;  
QY 46 TCTCTCAATGGCACTGCTATTATCCATTCGCAATACACATTCGAGTTCGAA 105  
DB 161 TTGTGTATTTGACAGAGCTCTTTTATACTAAGACAAATTTAATTTTGTACTAGAA 220  
QY 106 AAGGACTTCCCAAGTTATTGGAGTACTGTGAAGAGTTCGTCAAGATTACCCAAAG 165  
DB 221 AAAAATTTGAACATTTTAGTCTCTGTTATATAAATGTTAATTCAGAA--TTAGTTTAA 278  
QY 166 GACTTTACTATGTAATTAANTGTCAAACCTAGTACATCAATAAATTTCTAGTGG 225  
DB 279 GCCTTAATTAACATAATANTAGCTTTGGACACTTAAGAGCTCTAATTTCTGTGTA 338  
QY 226 CAAAAA 238  
DB 339 AAAAAA 351

RESULT 10

US-09-783-590-5790/c  
; Sequence 5790, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products  
; FILE REFERENCE: PO-16.2C1

CURRENT APPLICATION NUMBER: US/09/783,590  
CURRENT FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5790  
LENGTH: 499  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (9)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (127)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (211)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (233)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (262)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (270)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (302)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (320)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (330)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (337)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (345)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (346)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (348)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (368)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (375)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (387)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (388)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (391)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (394)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (395)  
OTHER INFORMATION: n equals a,t,g, or c

## RESULT 11

RESULT 14  
US-10-072-349-13/C

```

RESULT 15
US-10-398-221-3632
/ Sequence 3632, Application US/10398221
/ Publication No. US20040018514A1
/ GENERAL INFORMATION:
/ APPLICANT: KUNST, Frederik
/ APPLICANT: GLASER, Philippe
/ TITLE OF INVENTION: Listeria innocua, genome and applications
/ FILE REFERENCE: 344 702 - US
/ CURRENT APPLICATION NUMBER: US/10/398,221
/ CURRENT FILING DATE: 2003-03-27
/ PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
/ PRIOR FILING DATE: 2001-10-04
/ PRIOR APPLICATION NUMBER: FR 00/12 697
/ PRIOR FILING DATE: 2000-10-04
/ NUMBER OF SEQ ID NOS: 4025
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 3632
/ LENGTH: 3423
/ TYPE: DNA
/ ORGANISM: Listeria monocytogenes 4h

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3632

Query Match      13.8%; Score 32.8; DB 16; Length 3423;
Best Local Similarity 59.8%; Pred. No. 44;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 58 GCAACTGCTATTATCCATTCCGCAATCAATTTTCGGATGTTCTCGAAAGGACTTCCCA 117
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2377 GAATTGTTTTTTTATCTTTTCCCAAGCHCAAAACGAATGGTATGGATGAAACTTCGCT 2436
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 118 AAGTTATGGAGTACTGTGAAAGAGTTCCTCA 149
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 2437 CAACATCTGGCATTAGTACGAAATGTTTGGCA 2468
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: July 7, 2004, 16:28:39  
Job time : 345 secs